Improving Bicyclic Peptide Phage Display and Development of Sortase A Inhibitors

THÈSE N^{O} 6413 (2014)

PRÉSENTÉE LE 28 NOVEMBRE 2014

À LA FACULTÉ DES SCIENCES DE BASE
LABORATOIRE DE PROTÉINES ET PEPTIDES THÉRAPEUTIQUES
PROGRAMME DOCTORAL EN CHIMIE ET GÉNIE CHIMIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Inmaculada RENTERO REBOLLO

acceptée sur proposition du jury:

Dr R. Hovius, président du jury
Prof. C. Heinis, directeur de thèse
Prof. K. Johnsson, rapporteur
Prof. Ph. Moreillon, rapporteur
Prof. E. Tate, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2014

Acknowledgements

First and foremost, I would like to express my deepest appreciation to my advisor Professor Christian Heinis. Thank you for accepting me as part of the LPPT lab, for the guidance and scientific advice as well as for all the motivation and support during all these years, for letting me freedom to follow my ideas, for trusting me, for always encouraging my research. Doing a PhD under your supervision has been a great learning experience that I have enormously enjoyed.

I would also like to thank Prof. Philippe Moreillon, for allowing me to do part of the experimental work in his lab, for taking the time to listen to all my troubles, that were not few, for the advice and motivation during the lab meetings in your group, and finally for being part of my thesis jury. Thanks to Prof. Kai Johnsson, for also being part of this jury, the first-year exam jury, for the lab meetings together during the first months, and for forwarding my application to Christian in the first place! Thanks to Prof. Ed Tate for taking the trouble to travel to be part of the jury and to Dr. Ruud Hovious for chairing the committee.

Huge thanks to all past and present members of the group for making the lab such a wonderful place to work, for the nice fun helpful atmosphere. In particular, thanks Julia and Alessandro, *abuela* and *abuelo*. You welcomed me when I was still lost around the pipets, and taught me how scientist should think and work. If all scientists were like you, science would be a better world! I just hope I will manage to keep that spirit in the future. Thank you Lisa, we have worked back-to-back and side-to-side for four years and I have missed you every day since you left your desk! Thank you for all the times you turned around in your chair to ask me, and thank you for all the times you turned around to answer me, for reading my mind, for knowing what to say to cheer me up. Thanks also to Philippe, Silvia, Davide, Charlotte, Ranga, Khan, Shiyu, Ale, Camille, Sangram, Mike, Vanessa, Raphael, Jonas, Luca, Jeremy, Carly... thank you because there was not a single day I did not want to come to work. Thanks to the students that at one or other point of my PhD worked with me: Florian, Louise, Céline, Lorenz, Ismaël. Even if specific projects may not be part of this written "summary" of my thesis, your work has certainly contributed to it. I hope you learned and enjoyed working with me as much as I learned and enjoyed working with you.

Additional thanks go to the DMF group in UNIL, Jose, Greg, and particularly Shawna, with whom I had the chance to work more closely. For letting me work in your lab, all the help and advice with microbiology issues. I really appreciated your welcoming attitude, your interest in my project and that you were always willing to help. It was a pleasure to collaborate with such a great group!

Thanks to members of LIP, for all the times I have used their instruments, and for all the fun barbeque-madness at the lake. Nico and Rudolf, gym team! thanks for all the scientific and non-scientific chatting either doing push-ups or in front of a beer, and for the critical reading of this thesis! I feel very lucky to share a little bit of *almost* every evening with such smart scientists and fun friends as you are!

I would also like to mention Béatrice for all the administrative help, which made my life much easier in Switzerland; Anne Lene and people from the EDCH doctoral school and people of the core services in BCH, the magasin and the MS facility, it seemed to me that BCH was just a funnel of nice smily helpful people! Thanks to Keith Harshman from the HTS facility in UNIL, and Sara Martínez in CRAG-Barcelona, for all the troubleshooting advice with the Ion Torrent sequencing.

And out of the lab environment, this work would not have been possible without the continuous support of my family. Thanks mum and thanks dad, for making me a curious learner, for telling me tales about Galileo, for supporting every single decision I took in my life, even if I was getting further and further away, you were always there making my life easier. You are just awesome! when I will grow up, I want to be like you :). Thanks to my grandma too, for cooking always my favorite *tortillas* every time I come back!

Thanks to all my out-of-the-lab friends, thanks for keeping me *relatively* sane. Thanks to the Spanish in Lausanne: Sergio, Josele, César, Paloma, Xavi, Luis, Eva, Anna, Jorge, Paco, Javi, Ángela, little Màxim, Paul, Soraya... you brought all the Spanish-ness I needed to feel home. Thanks to the Spanish a bit further away, particularly Mer, Miwo and Cris. I know that my visits are less and less frequent, and that I keep calling the bars with the name they had seven years ago, but I can't help smiling when I'm walking towards *la moneda* to meet you, when you show me the newest place with yet a greater choice of ridiculously oversized free *tapas*, when we always end up in *La Ruina*. I miss you. Doctor, I could not survive in *la zona* without your updated playlists of *música decente* (sic), you keep Alcalá spirit in my car!

Tom, my final thanks go to you. First, scientifically. You have this perfect mix between the genuine curiosity of a little boy peeping into the fish brain, and the well-thought ambition of a researcher willing to take neurobiology one step further. This makes you one of the best scientists that I know, and you have definitely helped me to grow as one. And second, personally. You basically approach life the same way, as a fun challenge. It has been amazing to share all the fun challenges we have done these years, and I hope there are many more to come. Thanks for the scientific input, for the critical reading of all my manuscripts and of this thesis, for all the working weekends I kept you in the lab, for your continuous optimism, for carrying the heavy backpack in the mountains, for running at my pace when you could run faster. Thank you for being *una piedrita y la punta de un zapato*.

Lausanne, le 5 septembre 2014

Abstract

Bicyclic peptide ligands are promising molecules for the development of new therapeutics. They combine advantages from large protein therapeutics (high affinity and specificity) and small molecule drugs (accessibility to chemical synthesis and better diffusion into tissues). Large combinatorial libraries of bicyclic peptides can be generated and screened by phage display using a recently developed strategy. In brief, random peptides containing three cysteine residues are displayed on phage and reacted with a cyclization scaffold containing three thiol-reactive groups. The resulting bicyclic peptide library can be subjected to affinity selections following standard methods. Potent and selective bicyclic peptide inhibitors against several therapeutic targets have already been developed and their therapeutic potential is currently being evaluated in animal models.

One aim of my thesis was the exploration of ring size diversity in bicyclic peptides. Most selections had been performed with a library of uniformly sized peptides consisting of two loops of 6 residues each. I generated a set of libraries of the format Cys-$(Xaa)_m$-Cys-$(Xaa)_n$-Cys, where 'm' and 'n' = 3, 4, 5 or 6, and performed affinity selections against the tumor-associated serine protease urokinase-type plasminogen activator. Interestingly, bicyclic peptide inhibitors from virtually all ring size combinations were isolated, suggesting that many peptide formats can be accommodated in the active site of this enzyme. Moreover, they showed a large variety of consensus sequences and several of the identified consensus sequences were exclusively found in bicyclic peptides having specific ring size combinations. Some of these peptides may bind in orientations that allow affinity maturation of non-conserved regions, while others do not. Having available multiple leads isolated from such bicyclic peptide libraries with variable ring sizes could therefore be a great asset for the generation of high affinity binders. Additionally, other targets may have preferences for specific peptide constraints and the availability of these libraries increases the chances to isolate high affinity binders to any desired target.

A second goal of my thesis was to apply high throughput sequencing technologies to phage display selections of bicyclic peptides, in order to identify a larger number of specific target-binding sequences and motifs. I developed a procedure to efficiently compare the sequences of large numbers of phage-selected peptides to identify target-binding peptide motifs based on abundance and sequence similarity. Applying this approach to phage isolated in selections against five different protein targets, I was able to identify rare target-binding peptide motifs and could more precisely define groups and subgroups of consensus sequences. This information is valuable to choose peptide leads for drug development and facilitates the identification of epitopes. Moreover, binding motifs could be identified after a single round of phage panning. Such a selection regime reduces propagation-related bias and facilitates the application of phage display in non-

specialized laboratories, as procedures such as bacterial infection, phage propagation and purification are not required.

The final objective of my thesis was to discover bicyclic peptides that could be used as new antibiotics. Towards this end, I combined the newly generated variable ring size libraries and high-throughput sequencing procedures. I focused on the development of inhibitors of *Staphylococcus aureus* sortase A, an antivirulence target for which no potent and specific inhibitors have been reported. For the isolation of bicyclic peptide inhibitors to this target, the ring size diversity of the libraries turned out to be key. Inhibitors all shared the same motif (Leu-Pro-Pro) in a loop of 5 residues. Further characterization of their effects on *S. aureus* showed that they were not degraded by secreted proteases, and that they could inhibit sortase-mediated incorporation of external substrates on the staphylococcal cell wall. However, they were not sufficiently potent to compete with the native substrates of the enzyme, and therefore did not prevent their incorporation into the cell wall. More potent inhibitors are needed to effectively inhibit sortase A on *S. aureus* cells, and the bicyclic peptide inhibitors isolated constitute promising leads for the development of future antisortase therapeutics.

Keywords

peptide, bicyclic peptide, diversity, in vitro evolution, phage display, high-throughput sequencing, next generation sequencing, peptide antibiotics, antivirulence therapies, sortase A

Résumé

Les peptides bicycliques sont de prometteuses molécules pour le développement de nouveaux agents thérapeutiques. Elles combinent les avantages des protéines thérapeutiques de grande taille (caractérisées par leur grande affinité et spécificité pour leurs cibles) et des composés chimiques de taille plus réduite (ayant la faculté d'être synthétisés chimiquement et de mieux diffuser dans les tissus). De grandes librairies de peptides bicycliques peuvent être générées et criblées par "phage display" (expression phagique) en utilisant une stratégie récemment développée. En résumé, des peptides avec des séquences aléatoires contenant trois cystéines sont présentés à la surface de phages. Les groupements thiols agissent comme des nucléophiles qui vont réagir avec une molécule jouant le rôle d'armature du peptide bicyclique. Les phages contenus dans la librairie ainsi créée peuvent être soumis à une sélection par rapport à leur affinité moléculaire ce qui permet de mettre en évidence les séquences qui se lient aux cibles d'intérêt. Cette méthode a permis l'isolation d'inhibiteurs avec une grande efficacité et sélectivité contre plusieurs cibles thérapeutiques, et leur potentiel thérapeutique est à présent en train d'être évalué sur des modèles animaux.

Un des buts du travail présenté dans cette thèse a été d'explorer la diversité obtenue grâce à différentes longueurs des boucles dans les peptides bicycliques. La plupart des sélections avaient jusqu'alors été effectuées avec une librairie de peptides de taille constante (deux boucles de 6 résidus chacune). Ainsi, j'ai généré une collection de librairies au format Cys-$(Xaa)_m$-Cys-$(Xaa)_n$-Cys, où 'm' et 'n' = 3, 4, 5 ou 6, et j'ai effectué des sélections contre l'activateur du plasminogène de type urokinase, une protéase à serine impliquée dans le développement de certains cancers. J'ai pu isoler des peptides bicycliques inhibant la cible avec presque tous les formats. Cela suggère que des boucles de différente longueur peuvent adopter des conformations afin d'interagir avec le site actif de cette enzyme. Qui plus est, ces différents formats ont permis l'isolation d'une grande variété de séquences consensus et plusieurs d'entre elles étaient spécifiques d'un seul type de librairie particulière. Pour certains peptides, l'affinité peut être améliorée à cause de la présence de zones d'interactions plus faibles mais cela ne peut être généralisé. Par conséquent, la mise en évidence de multiples motifs peptidiques, provenant de librairies comportant différentes longueurs de boucles, représente un atout majeur pour la génération de ligands à forte affinité. En plus, certaines cibles peuvent présenter des préférences pour un format en particulier, et la disponibilité de plusieurs librairies augmente les chances d'isoler un ligand pour n'importe quelle cible désirée.

Le deuxième but de ce travail a été d'appliquer la technologie de séquençage à haut débit pour analyser les sélections de phages. Cela devrait permettre l'identification d'un nombre maximal de séquences et de motifs de liaison. J'ai développé une procédure pour comparer efficacement les séquences d'un grand nombre de peptides et pour identifier des motifs sur la base de leur abondance et de leur similarité. En utilisant cette approche

pour analyser les résultats obtenus avec des sélections contre cinq protéines cibles différentes, j'ai été capable d'identifier des motifs de liaison rares (qui auraient été omis avec la méthode classique) et de définir plus précisément les groupes et sous-groupes de chaque consensus. Cette information est d'une grande utilité pour choisir les candidats au développement des médicaments et elle facilite l'identification d'épitopes. En outre, cela a permis d'identifier des motifs de liaison après la première série de sélection phagique. Une telle procédure réduit les chances d'avoir des résultats faussés par les biais de propagation des phages, et facilite l'application du "phage display" en laboratoire non spécialisé, étant donné que les étapes telles que l'infection bactérienne, l'amplification et la purification des phages ne seront plus nécessaires.

Le but final a été de trouver peptides bicycliques qui pourraient être utilisés comme de nouveaux antibiotiques, en combinant les avantages des librairies avec différentes longueurs de boucles et ceux des procédures de séquençage à haut débit. Plus spécifiquement, je me suis concentré sur le développement d'inhibiteurs de la sortase A de *Staphylococcus aureus*, une cible impliquée dans la virulence des infections pour laquelle aucun inhibiteur puissant et spécifique n'a été rapporté. Pour l'isolation de peptides bicycliques inhibant cette cible, la diversité donnée par la longueur de boucle s'est révélée clé. Tous les inhibiteurs partageaient le même motif (Leu-Pro-Pro) dans une boucle à 5 résidus. Par la suite, la caractérisation de leurs effets sur *S. aureus* a montré qu'ils n'étaient pas dégradés par des protéases sécrétées, et qu'ils pouvaient inhiber l'incorporation de substrats externes sur la paroi cellulaire par la voie de la sortase. Cependant, ils n'étaient pas suffisamment puissants pour entrer en compétition avec les substrats naturels de cette enzyme, et par conséquent ils n'empêchaient pas l'ancrage de facteurs de virulence. Des inhibiteurs plus puissants sont requis pour inhiber efficacement la sortase A sur les cellules de *S. aureus*, et les peptides bicycliques isolés restent en bonne position pour le futur développement de médicaments antisortase.

Mots-clés

peptide, peptide bicyclique, diversité, évolution in vitro, expression phagique, séquençage à haut débit, séquençage de la prochaine génération, antibiotiques peptidiques, thérapies antivirulence, sortase A

Abbreviations

ACN	acetonitrile
AMP	antimicrobial peptide
ATP	adenosine-5'-triphosphate
BBMB	1,3-bis(bromomethyl)benzene
BSA	bovine serum albumin
CDR	complementarity determining region
DARPin	designed ankyrin repeat protein
DIPEA	N,N-diisopropylethylamine
DMMA	(Z)-3-(2,5-dimethoxyphenyl)-2-(4-methoxyphenyl) acrylonitrile
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDT	1,2-ethanedithiol
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
ESI	electrospray ionization
Fc region	fragment crystallizable region
FDA	Food and Drug Administration (US)
Fmoc	fluorenylmethyloxycarbonyl
Fn3	fibronectin type III domain
FXIIa	activated coagulation factor XII
HBTU	O-(benzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HOBt	1-hydroxybenzotriazole
HPLC	high-performance liquid chromatography
HTS	high-throughput sequencing
IPTG	isopropyl β-D-thiogalactopyranoside
LPS	lipopolysaccharide

K_d	dissociation constant
K_i	inhibition constant
mRNA	messenger ribonucleic acid
NMR	nuclear magnetic resonance
MALDI	matrix-assisted laser desorption ionization
MRSA	methicillin-resistant *Staphylococcus aureus*
MSCRAMM	microbial surface component recognizing adhesive matrix molecules
MS	mass spectrometry
NGS	next generation sequencing
NHS	N-hydroxysuccinimide
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	poly-ethylene glycol
PK	plasma kallikrein
PVS	phenyl vinyl sulfone
SA	streptavidin
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SrtA	sortase A
SrtB	sortase B
TATA	1,3,5-triacryloyl-1,3,5-triazinane
TBAB	*N*,*N*′,*N*′′-(benzene-1,3,5-triyl)-tris(2-bromoacetamide)
TBMB	1,3,5-tris(bromomethyl)benzene
TCEP	tris(2-carboxyethyl)phosphine
TFA	trifluoroacetic acid
uPA	urokinase-type plasminogen activator
VRSA	vancomycin-resistant *Staphylococcus aureus*

Contents

Chapter 1

Introduction

1.1 Objectives

The discovery of ligands with tailored binding properties to targets of interest is essential for the development of therapeutics and diagnostic tools, as well as for the study and manipulation of biological systems. In our laboratory, bicyclic peptide ligands against targets of interest are routinely developed using an affinity selection strategy based on phage display. A first objective of this work was to improve the performance of the selection process in two ways: (i) increasing the diversity of the bicyclic peptide libraries available by generating new libraries with different ring size combinations, and studying the influence of the ring size on the ligands identified; and (ii) maximize the amount of specific target-binding sequences and motifs obtained from phage display selections using high-throughput sequencing. The final aim was the application of the improved selection process to discover bicyclic peptides that could be used as new antibiotics. Specifically, I focused on inhibitors of *Staphylococcus aureus* sortase A, a potential antivirulence target.

1.2 Peptide therapeutics

Peptides are an attractive class of molecules for the development of novel therapeutics. Natural peptides are able to perform very specific and complex functions, and exhibit high specificity for their targets with minimal off-target effects[1,2]. One of the main strengths of peptides as drugs lies in powerful new approaches for the generation and screening of combinatorial peptide libraries[3], allowing peptide-based ligands to targets of choice to be evolved *in vitro*. These approaches include phage display and other selection technologies that will be described in the following section. Peptides can then be chemically synthesized and easily conjugated to desired functional groups, such as labels or tags. In therapy, their small size allows better tissue penetration than large protein therapeutics and reduces the risk of immune reactions[1].

However, peptides have been generally considered as poor drug candidates due to their low oral bioavailability, poor pharmacokinetics, limited systemic stability and poor membrane permeability[1,4]. In spite of the large diversities that can be generated and screened by *in vitro* selection technologies, very few of the identified candidates have reached the clinic. An example is peginesatide (Hematide®), a dimeric PEGylated erythropoietin-mimicking peptide, which was approved in 2012 by the FDA for the treatment of anemia associated with chronic kidney disease[5,6]. It was however withdrawn from the market soon after due to adverse side effects[7].

Almost all peptide drugs and drug candidates are naturally occurring peptides or derivatives thereof. In contrast to *in vitro* evolved peptides, most naturally occurring peptides have non-canonical structures which include macrocyclization and unnatural residues and linkages[8,9]. In particular, structural rigidity conferred by macrocyclization can offer several advantages in peptide ligands. First, the smaller entropic loss upon binding allows reaching higher target-binding affinities. Second, the lower number of possible conformers confers higher specificities. And third, structural constraints render peptides less accessible to protease cleavage, showing increased metabolic stability[10,11]. In certain cases, cyclization can additionally enhance membrane permeability[12].

There is therefore considerable interest in translating these non-canonical features into *in vitro* selection systems. Some of the approaches will be discussed in the following section, with special emphasis on the generation of phage display libraries of bicyclic peptides.

1.3 Phage display for the selection of peptide ligands

1.3.1 Phage display: technology overview

Phage display is a powerful technology for the isolation of protein or peptide ligands to targets of choice[13-15]. It is an *in vitro* selection system, pioneered by G. P. Smith in 1985, in which a library of polypeptide variants is expressed on the surface of phage particles as coat protein fusions[13]. Each particle displays a single polypeptide variant on the outside and contains the corresponding coding DNA inside. The linkage between phenotype (peptide displayed) and genotype (DNA coding for it) allows the isolation of ligands with specific binding properties through a process of affinity selection, and the identification of their primary sequence by DNA sequencing (Figure 1). Generally, several rounds of affinity selection, infection of bacteria and re-amplification of the phage are needed in order to enrich the population in target-specific binders[16,17].

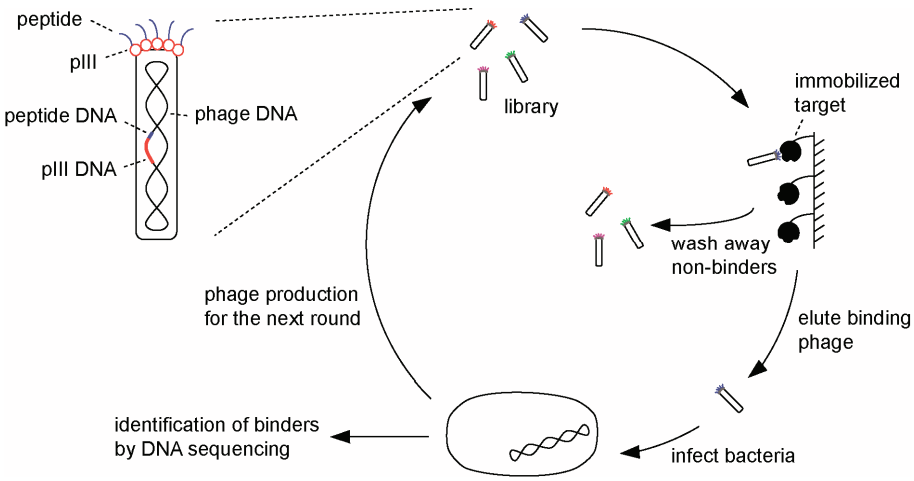


Figure 1. Scheme of a standard cycle of phage display selection. Combinatorial libraries of peptides are displayed on the surface of phage particles as coat protein fusions. In the picture, several copies of the peptide variant (blue) are displayed fused to the minor coat protein pIII (red). The phage library is incubated with an immobilized target. After extensive washings, target-bound phage are eluted and amplified in new host cells for a subsequent round of selection. After several cycles, peptide sequences are identified by DNA sequencing.

Filamentous phage strains M13 and fd are by far the most frequently used[17]. Phage particles of these strains consist of a rod-shaped protein coat enclosing the single stranded DNA viral genome (with 99% sequence identity between both strains). The coat proteins most frequently used for display are pVIII and pIII, although alternative systems on other coat proteins have also been described[15,17-20]. pVIII is present in more than 2000 copies and constitutes the body of the phage particle. pIII is present in five copies at one end of the phage particle and it is involved in bacterial infection. Different levels of valency (i.e. number of copies of the displayed peptide variant per phage particle) can be achieved by co-expressing wildtype coat proteins together with the peptide-displaying counterparts in order to produce a hybrid virion. Polyvalency (having multiple copies of the same variant on the phage) can be of advantage to isolate binders from naïve libraries, as it enhances capture of the phage even when the peptides have weak affinities. However, for affinity maturation purposes, monovalent systems might provide a better discrimination between high-affinity and low-affinity binders.

1.3.2 Protein and peptide ligands selected by phage display

The first molecules to be displayed on phage were linear polypeptides[13,21]. During the 1990s, Winter and co-workers applied the technology to the display of variable regions of antibodies[14]. The screening of phage antibody libraries has led to the development of several approved therapeutic drugs such as the blockbuster adalimumab, belimumab and raxibacumab[22-24], all three human monoclonal antibodies. In contrast, linear pep-

tides isolated by phage display had typically weaker affinities and have been mainly used for epitope mapping and research applications[25].

The generation of constrained peptide libraries by cyclization of linear peptides *via* a disulfide bridge between two cysteines allowed the isolation of peptides with higher affinities[26,27]. However, in most cases their affinity was still not sufficient for therapeutic applications (in the micromolar to high nanomolar range, in contrast to sub-nanomolar affinities of antibodies)[28]. Only one cyclic peptide isolated by phage display has reached the market: the previously mentioned peginesatide, a disulfide-cyclized erythropoietin mimetic, approved in 2012 and withdrawn soon after due to undesired side effects.

In addition to peptides and antibodies, phage display has recently been applied to a number of structurally diverse protein scaffolds based on the Z domain of protein A (affibodies), fibronectin domains (monobodies), lipocalins, DARPins, and many others[29-33]. Several candidates derived from these alternative scaffolds are currently under clinical development, such as the VEGFR2-targeting pegdinetanib[34] and the VEGF-targeting abicipar[35]. Both are currently in phase II clinical trials.

1.3.3 Chemical and enzymatic modifications of polypeptide phage display libraries

In order to explore structures or functionalities beyond the ones present in the 20 natural amino acids, peptides can be modified post-translationally on the surface of the phage[36,37]. In these cases, selectivity of the reaction is essential, as modifications in other coat proteins might reduce phage viability and infectivity.

Thiols from cysteine residues on the peptide constitute ideal reaction handles. Cysteine is one of the least abundant amino acids on phage proteins[36], and stable cysteine-free mutants of the pIII coat protein are available[38]. Thiol groups have been used to incorporate additional functionalities into linear polypeptides on phage, such as fluorophores or glycan moieties[39,40] (Figure 2A). In addition to disulfide bridge formation between two cysteine residues on the peptide, thiol groups have also been used for chemical cyclization of peptides on phage. A very recent example is the modification of phage libraries with photo-reactive scaffolds to generate light-responsive cyclic peptide ligands[41,42]. Phage libraries of peptides containing two cysteine residues were cyclized with a thiol-reactive azobenzene linker. Azobenzene undergoes a pronounced change in geometry upon UV irradiation from *trans* to *cis* conformation, which in turn changes the conformation of the peptide backbone (Figure 2B). Phage selection of azobenzene-cyclized peptide libraries led to the isolation of photoswitchable ligands with up to 4-fold difference in binding affinity between the *cis* and the *trans* isomers. Cysteine residues are also used for the generation of bicyclic peptides by chemical cyclization of phage-displayed linear peptides[43,44] (Figure 2C). This cyclization strategy was used in the present work and will be discussed in further detail in the following section.

Besides thiol groups, recently Derda and coworkers proposed aldehydes, obtained by periodate oxidation of N-terminal Ser/Thr residues, as unique reactive handles[45]. For example, using an aminooxy mannose derivative, they were able to generate glycopeptide libraries *via* oxime ligation (Figure 2D). More recently, sortase-mediated transpeptidation was also applied to incorporate proteins or small molecules on phage[46] (Figure 2E). Although cysteine residues still represent the most convenient reactive handle on phage, these new alternatives can be used to combine different modifications on the same peptide, further increasing the versatility and application of phage libraries.

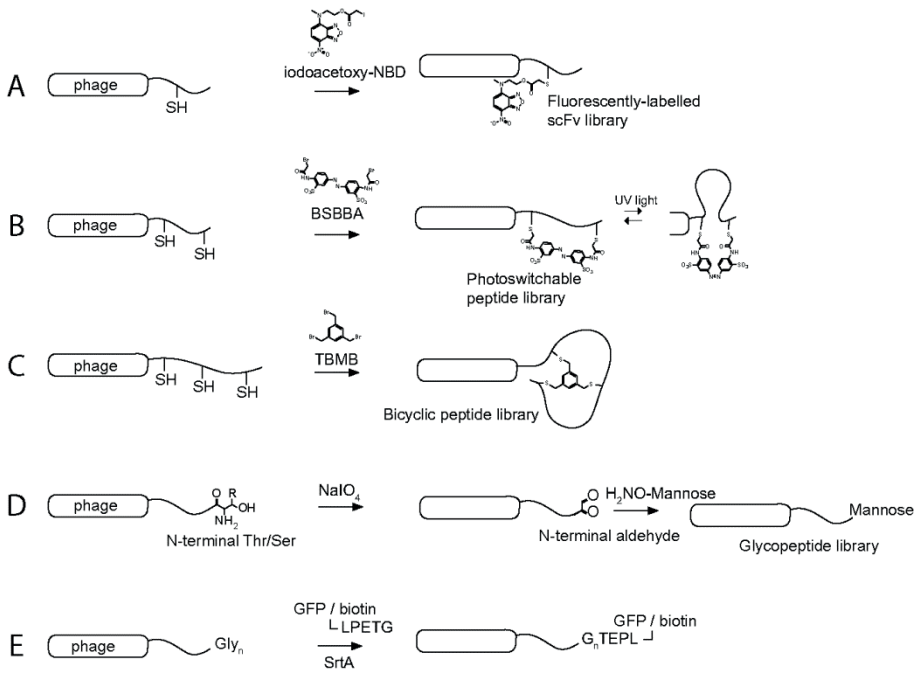


Figure 2. Modification of peptides on phage. Thiol groups represent convenient reactive groups for the incorporation of chemical groups (A) or for cyclization (B, C). Alternative strategies include the generation of N-terminal aldehydes (D) or the use of N-terminal polyglycine for SrtA-mediated transpeptidation (E).

1.3.4 Alternative *in vitro* selection technologies

Besides phage display, other *in vitro* selection technologies have been developed and applied for the isolation of peptidic ligands. They include mRNA display, ribosome display, yeast display, bacterial display, among many others. In contrast to phage display, mRNA display and ribosome display work entirely *in vitro*: in mRNA display the peptide variant is linked to the mRNA *via* puromycin[47], and in ribosome display *via* the ribosome[48]. These entirely *in vitro* systems have the advantage that a transformation step is not needed for the generation of the libraries, allowing larger libraries to be obtained with less effort. While the typical library size for phage display is 10^8 – 10^9 different variants, mRNA display and ribosome display libraries can reach 10^{14} different vari-

ants. Additionally, they are more flexible systems regarding the incorporation of unnatural amino acids. Although phage systems allowing the incorporation of non-natural residues have been described[49], they were limited to one additional amino acid. Additionally, the efficiency of incorporation of non-natural residues is lower than their natural counterparts, requiring the optimization of precise growth conditions to minimize this bias. mRNA display-based systems, such as the RaPID system developed by Suga and coworkers[8], can overcome some of these limitations. By using a custom-made cell-free translation system[50], arbitrary aminoacyl-tRNA synthetases can be omitted to leave the corresponding codons vacant. These are then assigned to an unnatural residue by adding the adequate aminoacyl-tRNAs prepared externally. Such tailor-made aminoacyl-tRNAs can be prepared using the flexizyme technology developed by the same group[51,52]. This versatile and flexible system has been used for the selection of nanomolar or even picomolar binders of Sirtuin2 (IC_{50} = 3.7 nM), VEGFR2 (K_d = 33 nM) and E6AP (K_d = 600 pM)[8]. However, the application of this system is technically complex and limits its widespread use by unspecialized laboratories.

Despite its limitations, phage display remains the most commonly used selection technology, probably due to its robustness, its versatility and, in the case of peptides, the availability of commercial libraries.

1.4 Previous work on bicyclic peptides

1.4.1 Natural bicyclic and multicyclic peptides

Bicyclic and multicyclic peptides are produced by organisms of all kingdoms, both by ribosomal and non-ribosomal synthesis, and have very diverse biological activities. Cyclization linkages in ribosomally synthesized peptides are typically disulfide, amide, thioether or ester bonds. For example, sunflower trypsin inhibitor 1 (SFTI-1) is a bicyclic peptide cyclized head-to-tail and containing an internal disulfide bridge (Figure 3)[53]. Similarly, plant cyclotides are multicyclic peptides cyclized head-to-tail and bridged by two or more disulfide bonds[54]. Vertebrate defensins are antimicrobial peptides cyclized by three intramolecular disulfide bridges, where the subfamily of theta-defensins are also cyclized head-to-tail[55]. Multicyclic peptides produced by nonribosomal synthesis typically contain unnatural amino acids and show a greater variety of cyclization chemistries. Two examples are the marketed bicyclic peptide drugs actinomycin-D, an antibiotic and chemotherapy agent that interferes with mRNA synthesis, and romidepsin, an anticancer agent inhibiting histone deacetylases (Figure 3).

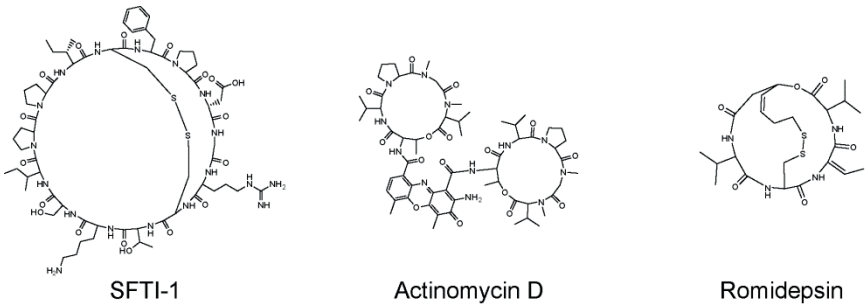


Figure 3. Examples of different cyclization chemistries in natural bicyclic peptides. The ribosomally synthesized SFTI-1 is cyclized head-to-tail and contains an internal disulfide bridge. Actimonycin D and Romidepsin are non-ribosomally synthesized. Actinomycin D contains two identical lactone rings connected through an aminophenoxazin group. Romidepsin is a cyclic depsipeptide with an internal disulfide bridge. This disulfide bridge undergoes reduction within the cells and binds to the Zn atom in the active site of histone deacetylases.

1.4.2 Phage selection of bicyclic peptides

Combinatorial libraries of bicyclic peptides can be generated by chemical cyclization of phage-displayed linear peptides *via* cysteine residues. In a proof-of-concept work in 2009, bicyclic peptide libraries were generated by reacting linear peptides containing three cysteine residues with the thiol-reactive cyclization scaffold 1,3,5-tris-(bromomethyl)benzene (TBMB) (Figure 4A). TBMB and other bromomethylbenzene derivatives had previously been proposed as suitable reagents for quantitative cyclization of peptides in aqueous solutions (CLIPS™ technology[56,57]), and such reactions proved compatible with phage. Selections of TBMB-modified phage libraries against two human serine proteases, plasma kallikrein and cathepsin G, allowed the isolation of potent bicyclic peptide inhibitors[43] (Figure 5). Since then, this system has been applied to a variety of targets, resulting in the isolation of inhibitors and ligands with high binding affinities (typically in the nanomolar range) and specificities[44,58,59].

In this system, two elements can be varied to increase the diversity of the libraries: the cyclization scaffold and the peptide loop lengths. Our laboratory recently developed three new cyclization linkers for the generation of phage libraries of bicyclic peptides (Figure 4B), namely 1,3,5-triacryloyl-1,3,5-triazinane (TATA), *N*,*N*′,*N*′′(benzene-1,3,5-triyl)-tris(2-bromoacetamide) (TBAB), and *N*,*N*′,*N*′′-benzene-1,3,5-triyltrisprop-2-enamide (TAAB)[60]. They were designed to provide different environments to the peptides, presenting different geometries and having additional hydrogen bond donors and acceptors that could establish further interactions with the backbone or side chains of the peptide loops. Two of the linkers, TATA and TBAB, have been successfully applied to phage selections. Moreover, the crystal structure of one of the TBAB-selected bicyclic peptides showed that this scaffold was able to establish hydrogen bonds with the backbone and side chains of the selected peptide[61]. These new scaffolds can be combined with

existing libraries and therefore triplicate their diversity. However, since they are not genetically encoded, selections with different scaffolds must be performed separately and competition among them is not possible.

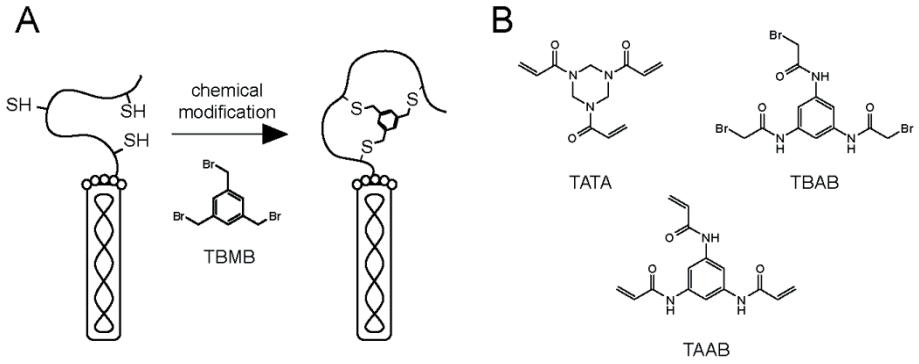


Figure 4. Generation of phage display libraries of bicyclic peptides. (A) Phage display libraries of peptides containing three cysteine residues are chemically modified with a thiol-reactive scaffold, resulting in combinatorial libraries of bicyclic peptides. TBMB was the first scaffold to be used and provides a hydrophobic core. (B) New scaffolds developed to cyclize peptides on phage. They contain hydrogen-bond donor and acceptor groups that can further establish interactions with the peptide.

Concerning loop lengths, most selections performed so far used the so-called 6×6 library, containing two loops of six amino acids each. Libraries of shorter bicyclic peptides (3×3 and 5×5, containing two loops of three or five amino acids each, respectively) were used in selections against human plasma kallikrein (PK)[62]. In those selections, the specificity profiles of the inhibitors could be tuned by modulating the size of the macrocyclic rings. For example, the 6×6 inhibitor PK15 inhibited human and monkey PK in the low nanomolar range, rat PK in the micromolar range, and did not inhibit any of the paralogous proteases tested (such as human factor XIa, thrombin, plasmin or factor XIIa). 5×5 inhibitors inhibited human, monkey and rat PK in the nanomolar range, but not paralogous proteases. Shorter bicyclic peptides of the 3×3 format inhibited human, monkey and rat PK as well as human factor XIa in the nanomolar range. For this target, the 5×5 inhibitors were therefore best suited for drug development as they inhibited orthologous proteases (allowing their use in animal models) but not paralogous proteases (which would lead to undesired side effects).

Recently, our group also demonstrated that, when libraries with three cysteine residues are subjected to panning without chemical modification, peptides with a fourth cysteine are strongly enriched[63]. Most probably, unpaired cysteines form disulfide bridges with cysteines of neighboring peptides, thus impairing infection and causing a less efficient propagation of such clones. This limitation of phage systems turned out to be of advantage for the screening of many different peptide topologies. Starting from libraries of peptides containing three cysteine residues, the fourth cysteine can appear in any of the

randomized amino acid positions, allowing the generation of a large number of topologically diverse bicyclic structures. In selections against two different model targets, a variety of binders with different topologies could be identified (Figure 5).

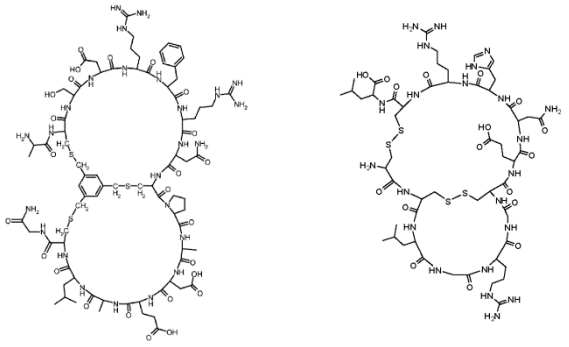


Figure 5. Examples of bicyclic peptides selected from phage combinatorial libraries. PK15 was identified from a phage library of TBMB-cyclized bicyclic peptides[43], and it is a potent inhibitor of human plasma kallikrein (K_i = 2 nM). UK504 was identified from a phage library of bicyclic peptides cyclized *via* two disulfide bridges, and inhibits human urokinase-type plasminogen activator (K_i = 7.7 μM).

1.4.3 Phage selected bicyclic peptides

Phage selection of bicyclic peptides has been applied to a variety of targets, including plasma kallikrein[43], cathepsin G[43], urokinase-type plasminogen activator (uPA)[44], coagulation factor XII[58], proteases of other classes (unpublished data) and Her2 receptor[59].

One of the best characterized bicyclic peptide ligands is the uPA-inhibitor UK18 (Ac-A<u>C</u>SRYEVD<u>C</u>RGRGSA<u>C</u>G-NH_2, cyclized *via* reaction of TBMB with the underlined cysteine residues, Figure 6). Its target, uPA, is a trypsin-like serine protease involved in tumor growth and migration[64]. It is a key component of the proteolytic cascade leading to active proteases responsible for the degradation of the extracellular matrix and other biological barriers. UK18 was originally identified from a 6×6 bicyclic peptide library modified with TBMB[44]. It is a potent and selective inhibitor, with a K_i of 53 nM for human uPA and much weaker inhibitory activities (K_is of 111-316 μM) for related proteases. Based on the X-ray structure of the UK18-uPA complex, a more potent UK18 derivative (UK202, K_i = 28 nM) was recently developed by replacement of the glycine at position 13 with D-serine[65].

Another promising bicyclic peptide is the FXII-inhibitor FXII402 (Ac-G<u>C</u>GGRP<u>C</u>PPAY<u>C</u>G-NH_2, cyclized *via* reaction of TBMB with the underlined cysteine residues), with a K_i of 1.2 μM and more than 100-fold selectivity towards related proteases[58]. It was identified by selection of a 4×4 bicyclic peptide library cyclized with TBMB, followed by affinity maturation by rational design. Potent and selective inhibitors of FXII constitute valu-

able tools in hematology research and are attractive clinical candidates for antithrombotic therapy.

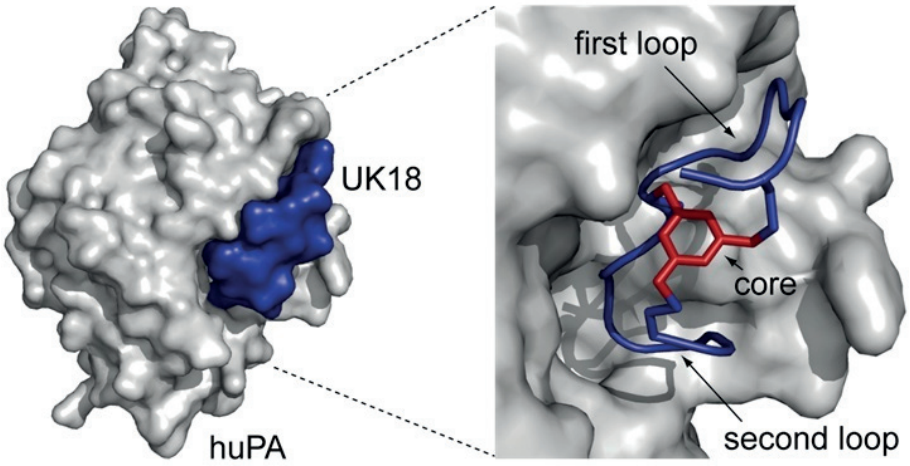


Figure 6. Crystal structure of bicyclic peptide UK18 bound to its target uPA. Both loops are involved in the binding, establishing a large interaction surface of 701 Å^2. Reprinted with permission from Angelini *et al*. ACS Chem. Biol[44]. Copyright 2012 American Chemical Society.

1.4.4 *In vivo* studies with phage-selected bicyclic peptides

Despite successful examples of peptide drugs available on the market, there are several limitations for the application of peptides as therapeutics, such as susceptibility to protease cleavage in plasma, short half-life due to renal clearance, low oral bioavailability and poor membrane permeability.

Further work on the bicyclic peptide UK18 described above showed that UK18 was more resistant to proteolysis in plasma *ex vivo* than its monocyclic and linear counterparts, or the two separate rings alone[66]. The rigidity of the rings, the smaller size and reciprocal protection to a certain extent seemed to cause this increased stability. Moreover, substitution of certain amino acids with unnatural residues could confer additional improvements. For example, the UK18 derivative UK202, in which a glycine was substituted by D-serine, showed 4-fold higher proteolytic stability[65].

Concerning renal clearance, pharmacokinetic studies with UK18 showed an elimination half-life of 30 minutes. In order to reduce renal clearance, two strategies have been explored: conjugation with an antibody Fc fragment[67] and conjugation with an albumin-binding peptide[66] (SA21, developed by Denner and coworkers[68]). In both cases half lives could be substantially increased (36 h for the Fc-conjugate and 24 h for the SA21-conjugate). More recent work has demonstrated that SA21-conjugated bicyclic peptides efficiently diffuse into organ tissues and solid tumor[69].

Low oral bioavailability and poor membrane permeability still constitute limitations for the application of bicyclic peptides as therapeutics. Bicyclic peptides would need to be administered intravenously, like most peptide drugs, and their application might be limited to extracellular targets. These represent however minor limitations and leave a wide range of pathological processes or targets accessible to bicyclic peptide drugs.

In summary, bicyclic peptides represent a promising molecule format for the development of new therapeutics. The strategy to generate and screen large combinatorial libraries of bicyclic peptides was developed only recently and it has already shown very promising results. Ongoing work is focused both on improvements for *in vivo* therapeutic application of several bicyclic peptides leads, as well as on the discovery of new leads against different targets.

1.5 High throughput sequencing (HTS) of DNA-encoded libraries

One of the strengths of *in vitro* selection methods is the easy identification of the selected peptides by DNA sequencing. Over the last 30 years, DNA sequencing technology has been dominated by Sanger sequencing and fluorescence-based electrophoresis[70,71]. Capillary-based, semi-automated implementations of Sanger biochemistry have been routinely used by virtually every biological and biomedical laboratory in basic and applied research, in academia, industry and clinic. It has led to a number of accomplishments, by providing a tool for decoding complete genes and later genomes[72]. In the last decade, several high throughput sequencing (HTS) technologies – also called next generation sequencing (NGS) technologies – have emerged. They are able to provide unprecedented volumes of sequence data in a rapid and economic manner[73], and have revolutionized the field of DNA sequencing and genomics.

1.5.1 HTS technologies

At present, multiple platforms coexist in the market, with each having advantages for specific applications over others. Rather than individual clones, HTS platforms sequence whole libraries of DNA fragments, where the total number of fragments that can be sequenced and their maximum length depend on the platform. A summary of the characteristic of the major HTS platform families is given in Table 1. Three technologies have been mainly applied for the sequencing of DNA-encoded libraries, and will be described in this section: Roche 454 pyrosequencing, Illumina dye sequencing and Ion semiconductor sequencing (also known as Ion Torrent). The sequencing process can be divided in three steps for all three technologies: (i) library preparation, (ii) clonal amplification of library fragments, and (iii) the sequencing reaction itself.

The library preparation consists on the generation of a library of DNA fragments of the adequate length flanked by platform-specific adaptors. For genome sequencing, it is necessary to shear the DNA in shorter fragments and to ligate them to the adaptor sequences. In the case of DNA-encoded libraries, DNA fragments can be amplified by PCR using suitable primers containing the adaptor sequences.

Next, library fragments are clonally amplified in localized *foci*. This step is necessary to increase the signal that will be detected during the sequencing reaction. 454 pyrosequencing and Ion Torrent use emulsion PCR to amplify library fragments on the surface of beads, each bead bearing a single template. In the Illumina system, library fragments are amplified by "bridge amplification" to form clusters of identical clones *in situ* on the surface of the flow cell that will be used for the sequencing reaction.

Platform family	Clonal amplification	Sequencing reaction	Max. read length	Max. # of reads
454	Emulsion PCR	Pyrosequencing (seq-by-synthesis)	700 bp	1 Million
Illumina	Bridge amplification	Reversible dye terminator (seq-by-synthesis)	300 bp 150bp	25 Million 3 Billion
Ion Torrent Ion Proton	Emulsion PCR	Proton detection (seq-by-synthesis)	400 bp 200 bp	5 Million 80 Million
SOLiD	Emulsion PCR	Oligonucleotide ligation (seq-by-ligation)	75 bp	3 Billion
PacBio	N/A (single molecule)	Fluorescent nucleotides (seq-by-synthesis)	8500 bp	0.8 Million

Table 1. Summary of the major NGS platforms available.

Finally, the sequencing reaction relies in the three cases on the so-called "sequencing by synthesis": a polymerase is used to synthesize the complementary strand of a single-stranded template, and the sequential incorporation of nucleotides is monitored. In the case of 454 pyrosequencing, each bead is placed on a different microwell on a chip, together with other smaller beads bearing the necessary enzymes (ATP-sulfurilase and luciferase). In each cycle, a single species of unlabeled nucleotide is introduced and incorporation events are detected thanks to pyrophosphate (PPi) release. ATP sulfurylase converts PPi to ATP, which allows the luciferase-mediated conversion of luciferin to oxyluciferin, generating visible light. The amount of light is proportional to the ATP and therefore to the number of nucleotides incorporated, and can be detected by a camera. Ion Torrent works similarly, but detects the release of protons upon nucleotide incorporation through the change of pH, which is also proportional to the number of nucleotides incorporated. In this case no additional enzymes or substrates are needed. Illumina sequencers rely on reversible dye terminators. All four nucleotides are provided in each cycle (carrying different fluorescent labels and with a 3'-OH blocking group) and the corresponding one is incorporated by the polymerase. After washing, the flow cell is im-

aged and then the fluorescent group is cleaved leaving the 3'-OH deprotected for the next cycle.

1.5.2 Application of HTS to *in vitro* selections

High throughput sequencing for the analysis of *in vitro* selection processes is becoming increasingly popular. The first technology to be developed was 454 pyrosequencing, and was also the first one to be applied to selections of DNA-encoded libraries[74]. Illumina sequencing has been the most extensively used technology, and was employed for the characterization of phage libraries of peptides and antibodies[75,76], mRNA libraries of fibronectin domains[77] and ribosome display libraries[78]. There are also some examples of the application of Ion Torrent for phage peptide libraries and mRNA peptide libraries[79,80].

For the sequencing of antibody repertoires, the long sequences provided by 454 pyrosequencing (Table 1) are well suited for V region analysis[81]. Illumina has also proven especially useful for phage antibody screenings. It provides a significantly higher throughput (> 100-fold), although the shorter read length has restricted its application for the analysis of CDR3 regions. The CDR3 regions of both light and heavy chains have proven to be the most relevant contributors to antibody specificity, and in some designed libraries are the only regions to be randomized[76]. For sequencing of peptide libraries, where short reads do not represent a limitation, Illumina sequencers have been the most commonly used[75,82].

HTS technologies have been mainly used for genome sequencing and for large-scale comparative and evolutionary studies. For such applications, the existence of genome assemblies of most model organisms has boosted the use of these short-gun technologies, which provide a large number of short reads. Most software therefore focuses on genome mapping of sequenced fragments to existing assemblies. In the case of *in vitro* selection technologies, HTS allows to get an unprecedented coverage of the selection outputs, which before was limited to a maximum of a few hundred clones. However, the bioinformatic challenges are very different and there are not yet broadly applicable tools to analyze HTS results from *in vitro* selections. As a consequence, although phage display is a wide spread technique used in many laboratories (especially due to the availability of commercial libraries), very few laboratories have the capabilities to apply HTS to characterize their selection results.

1.5.3 Ion Torrent sequencing

In this work, the Ion Torrent PGM platform was chosen for the analysis of phage selections, since it provides a suitable read length and a good trade-off between throughput and cost. Using the Ion 316™ Chip, more than 3 million sequences can be obtained. This is enough to cover the diversity of standard outputs from phage selections (typically between 10^3 and 10^5), and even allows multiplexing several samples in the same chip,

further reducing the costs. The workflow for sample preparation and sequencing is depicted in Figure 7. In brief, the region coding for the bicyclic peptide is amplified from the phage vector using suitable primers, containing adaptor sequences (needed for the subsequent steps) and a barcode. The barcode is a short sequence tag, typically 4-6 nucleotides, that allows the parallel sequencing of several samples in the same chip (multiplexing). The resulting PCR products are then clonally amplified on the surface of beads by emulsion PCR. Individual DNA molecules and primer-coated beads are isolated in aqueous droplets within an oil phase, where each droplet constitutes a PCR microreactor that amplifies a single DNA template. To guarantee only one template molecule per droplet (or none, but not more than one), most of the emulsion will be by definition empty of template molecules. After breaking the emulsion, the PCR-positive beads are separated from the non-templated beads, and loaded on the sequencing chip. The chip consists of millions of independent wells that essentially act as pH-meters, monitoring nucleotide incorporation by measuring the direct release of protons. The signal is converted to a sequence of nucleotides, generating two files: a "standard flowgram format" (SFF)- formatted file, which contains information about nucleotide flows that both did and did not result in base incorporation; and a FASTQ-formatted file, which contains the resulting sequence with quality scores. The Ion Torrent system also includes a pre-processing step, consisting of: (i) quality checks, to remove reads resulting from mixed DNA templates and/or having low signal quality, and (ii) trimming of 3' adaptor sequences.

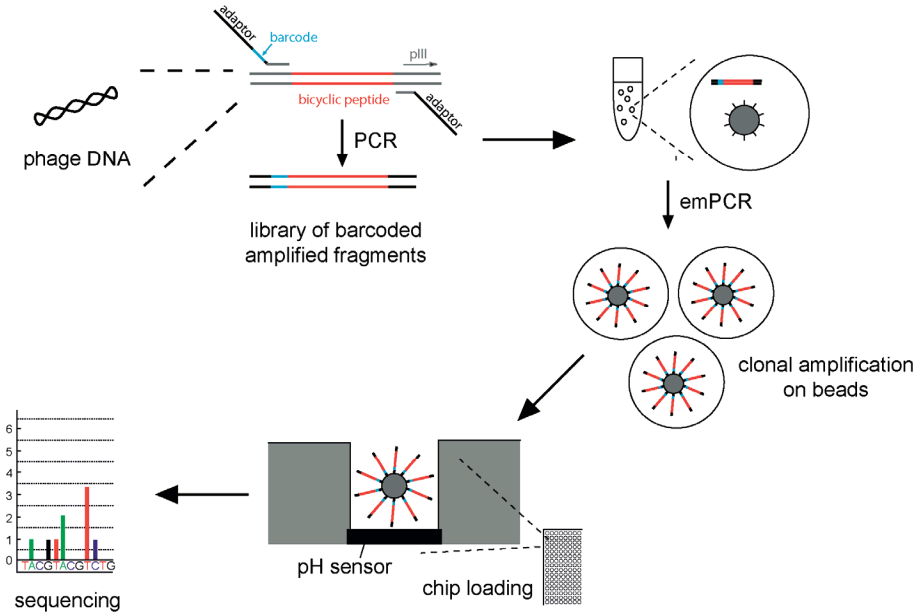


Figure 7. Ion Torrent sample preparation and sequencing workflow. First, the region of interest is amplified by PCR using primers containing adaptor sequences and barcodes. Then, library fragments are clonally amplified on beads by emulsion PCR. DNA-bearing beads are loaded on the sequencing chip, containing millions of independent wells. For the sequencing reaction, nucleotides are added sequentially and incorporation events are monitored by pH changes due to proton release.

1.6 Peptides as antibacterial agents

The World Health Organization's 2014 report on antimicrobial resistance highlights that antibiotic resistance is putting at risk the ability to treat common infections in the community and hospitals[83]. The report draws special attention to untreatable gonococcal infections resistant to third generation cephalosporins, multidrug resistant *E. coli* and *K. pneumoniae* infections, and methicillin-resistant *S. aureus* (MRSA). Despite significant efforts by pharmaceutical companies, very few antibiotics have reached the clinic in the last decades, and most are derivatives of previously approved drugs[84,85]. There is therefore a need of both new antibiotics and antibiotic-discovery platforms able to tackle this inevitable emergence of resistance. Phage libraries of bicyclic peptides could constitute a source of diversity to screen for new antibacterial compounds.

1.6.1 Natural peptide antibiotics

Many natural antibiotics and antimicrobial compounds are based on peptides. There are examples of both non-ribosomally synthesized (e.g. gramicidins, polymyxins, glycopeptides) and ribosomally synthesized (e.g. human defensins and bacteriocins) peptide antibiotics. Their mechanisms of action are also very diverse. The term "antimicrobial peptides" (AMP) is typically used to refer to short cationic peptides present throughout all kingdoms of life as part of the innate immune defenses[86]. They disrupt bacterial membranes causing cell death, although the precise mechanism is still a subject of debate[87]. Other peptide antibiotics exert their action through binding to specific targets. For example, vancomycin binds to the D-Ala-D-Ala moiety of lipid II cell wall precursors, thus preventing their use for cell wall biosynthesis. Nisin has a dual mechanism of action: the N-terminal region binds to the phosphate groups in lipid II, preventing cell wall biosynthesis, and the C-terminal region form pores on the membrane[88]. Some peptide antibiotics are even able to target intracellular processes, like the thiopeptide thiostrepton, which blocks translation by interfering with the elongation factor G in the ribosome[89].

1.6.2 *In vitro* evolution of peptide-based antibiotics

In vitro selection technologies have also been applied to find ligands to antibacterial targets with the aim of developing new drugs, although with limited success. One of the few successful examples is raxibacumab, a human monoclonal antibody for the treatment of inhaled anthrax[24], originally isolated from a phage display library. It specifically binds the protective antigen (PA) protein, a component of anthrax toxin, and prevents its binding to the anthrax toxin receptor with an IC_{50} of 0.5 nM. It does not directly harm the pathogen but acts by preventing toxin-mediated damage to the host.

Peptide libraries have been screened for inhibitors of several antibacterial targets such as Mur enzymes, involved in cell wall synthesis, β-lactamases, responsible for β-lactam antibiotic resistance, and other proteins[85,90,91]. However, only weakly active peptides

have been identified. An exception is the beta-hairpin peptide POL7080, a drug candidate against *Pseudomonas* infections[92]. It was discovered through iterative optimizations from a synthetic library of cyclic peptidomimetics that were designed based on the cationic antibacterial peptide protegrin-1, and cyclized through the beta-hairpin promoting di-peptide D-Pro-L-Pro. It showed potent activity against *Pseudomonas* spp. (including drug-resistant strains), while being inactive against other species of gram-negative and gram-positive bacteria. The target LptD, an outer membrane protein that transports LPS, was identified in a subsequent genetic screen and was a previously unexploited antibacterial target. POL7080 recently completed Phase I clinical trials.

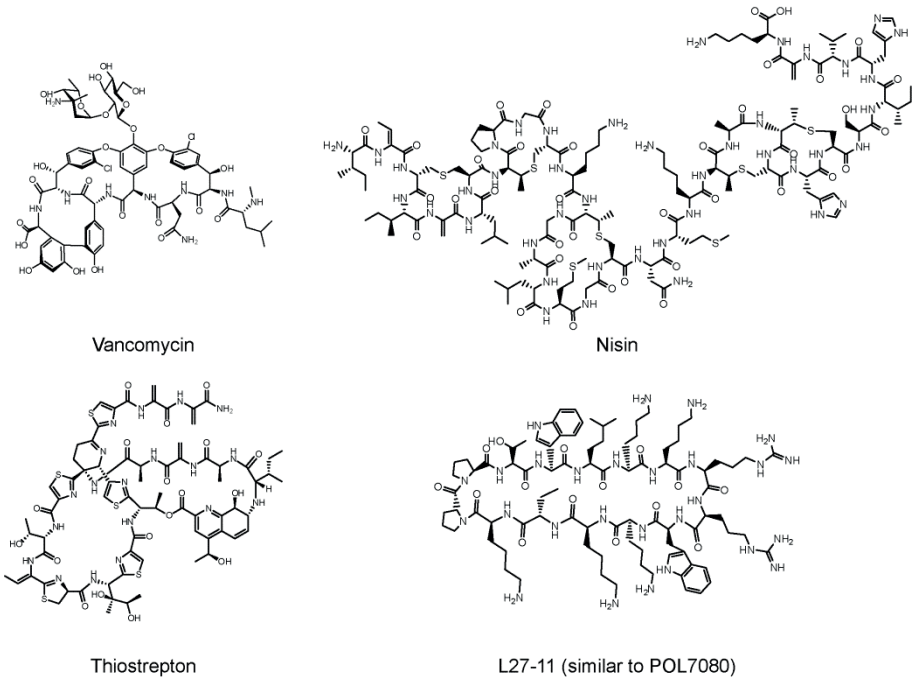


Figure 8. Examples of peptide antibiotics.

1.6.3 Targeting *S. aureus* pathogenesis

S. aureus is one of the main causes of fatal infections. It causes both superficial and invasive infections, including toxic shock syndrome, pneumonia, sepsis and infective endocarditis. Antibiotic treatment is often ineffective due to the development of resistant strains. Most isolates are penicillin resistant (between 65% and 90% in an European study on commensal *S. aureus* in 2013)[93]. Moreover, the prevalence of methicillin-resistant *S. aureus* (MRSA) strains in hospitals is alarming, and community-acquired MRSA cases are increasingly common[94]. The last-resort antibiotic for the treatment of MRSA strains is vancomycin, but vancomycin-resistant strains (VRSA) have also been reported[95]. Although VRSA strains are fortunately still rare, there is an urging need to discover new effective *S. aureus* drugs.

S. aureus pathogenesis can be divided in 5 stages: (i) colonization, (ii) local infection, (iii) systemic dissemination and sepsis, (iv) metastatic infection, and (v) toxinosis. Between 20% and 75% of the population is permanently or transiently colonized by *S. aureus*, which is carried asymptomatically. It behaves as an opportunistic pathogen, initiating infections when damage in skin or a mucosal barrier allows access to tissues or the bloodstream. Once in the vascular system, staphylococci can adhere to endothelial cells and be phagocytized. This intracellular environment can further protect the pathogen from the immune system and antibiotics, causing recurrent infections. It can also spread and colonize distant organs. Septic shock can ensue due to systemic infections or from locally secreted toxins, even when the bacteria have not reached the blood.

There are several factors contributing to the success of *S. aureus* as a human pathogen: the abundance of cell-surface and secreted virulence factors[96], and the great genetic repertoire for adapting to hostile environments[97]. In the search for new antibacterial approaches, antivirulence therapies have been proposed as potential alternatives. They aim at reducing virulence (pathogen-induced host damage) without killing the pathogen, offering a reduced selection pressure for drug-resistant mutations and preserving the host endogenous microbiome. Phosphosulphonates, for example, are drugs that inhibit *S. aureus* enzyme CrtM, responsible for the synthesis of a pigment that protects the bacteria from reactive oxygen species (ROS) produced by phagocytic defense of the host. *S. aureus* are not directly harmed by the drug, but are more susceptible to be killed by ROS at the site of infection. Proposed antivirulence strategies also include preventing toxin-mediated damage (as in the case of the previously mentioned raxibacumab, pg. 15), inhibiting "quorum sensing" (chemical signaling among bacteria that trigger responses to cell density and environment changes), and impeding microbial attachment and invasion of host tissues[98,99]. Regarding this last strategy, sortases, the enzymes responsible for the anchoring of adhesins and other virulence factors to the cell wall, have emerged as promising targets and will be described in the next section.

1.6.4 Sortase A as an antivirulence target

Sortases are transpeptidase enzymes from gram positive bacteria that anchor secreted proteins to bacterial cell surfaces (Figure 9). They specifically cleave within a sorting motif, present at the C-terminus of surface proteins, and covalently anchor the resulting carboxyl end to the cell wall. They are classified into different groups based on sequence homology, the substrate motif they recognize, and the acceptor nucleophile of the transpeptidation reaction. In *S. aureus* two sortase enzymes have been described: sortase A (SrtA, recognizing the motif LPXTG), responsible for the anchoring of proteins involved in adhesion to host cells and tissues and in immune evasion[100]; and sortase B (SrtB, recognizing the motif $NP^{Q}/_{K}TN$), which anchors proteins involved in iron uptake[101]. Interest in sortase as an antivirulence target comes from findings that loss of SrtA in *S. aureus* led to reduced pathogenicity. In a kidney abscess model, SrtA knockout mutants showed a 2-log decrease in cell titers in kidney lesions, and a 1.5-log increase in

LD_{50} (dose causing death in 50% of the mice)[102]. Similar results were observed in mouse models for infective endocarditis, septic arthritis and lung infection[103-106]. In contrast, SrtB contributed only marginally to the pathogenesis of staphylococcal infections, although it seemed to be involved in the persistence of the infection[103].

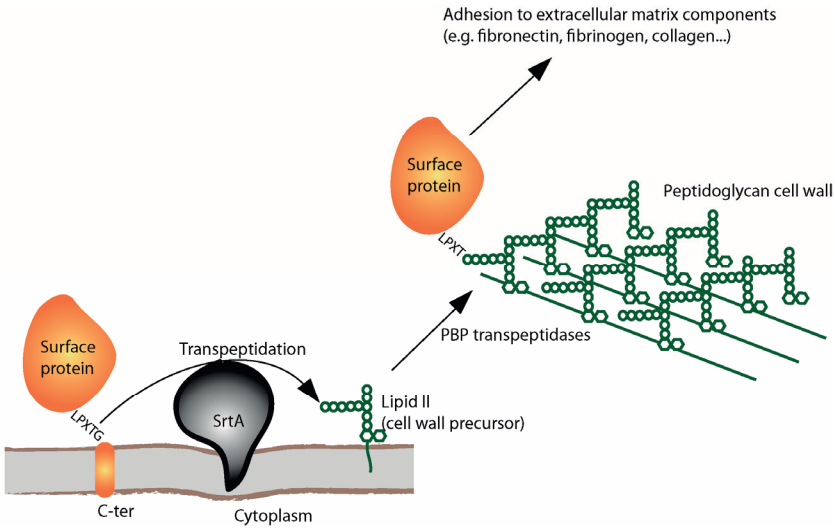


Figure 9. SrtA recognizes membrane-bound precursors of surface proteins by the conserved LPXTG motif and covalently anchors them to the pentaglycine moiety of lipid II.

SrtA from *S. aureus* is a 206-residue cysteine protease with an N-terminal membrane anchor. Analysis of the NMR structure revealed a previously unseen β -barrel structure with eight β-strands (Figure 10B and C), with the active site within an elongated hydrophobic groove. A histidine residue at position 120 is located in close proximity to the catalytic cysteine residue (position 184). Both residues are essential for SrtA activity[107], and most evidence suggest a reverse protonation catalytic mechanism, where only a minor fraction of the enzyme is catalytically active[108] (Figure 10A). It cleaves surface protein precursors between the threonine and the glycine of the LPXTG motif, forming a thioester bond with its active site sulfhydryl. Nucleophilic attack of the amino terminus of the pentaglycine in lipid II peptidogylcan precursor completes the sorting reaction.

The repertoire of cell wall-anchored proteins varies among strains. SrtA-anchored proteins include MSCRAMMs (microbial surface components recognizing adhesive matrix molecules), which drive adhesion to fibrinogen, fibronectin, collagen and other matrix molecules in the host; and protein A, which binds the Fc portion of immunoglobulins and is therefore involved in immune system evasion. Surface proteins show functional redundancy (e.g. at least five cell wall anchored proteins bind the plasma glycoprotein fibrinogen). Knockout mutants of single proteins are only partially defective in the studied function, thus the interest in targeting the common SrtA-mediated anchoring step.

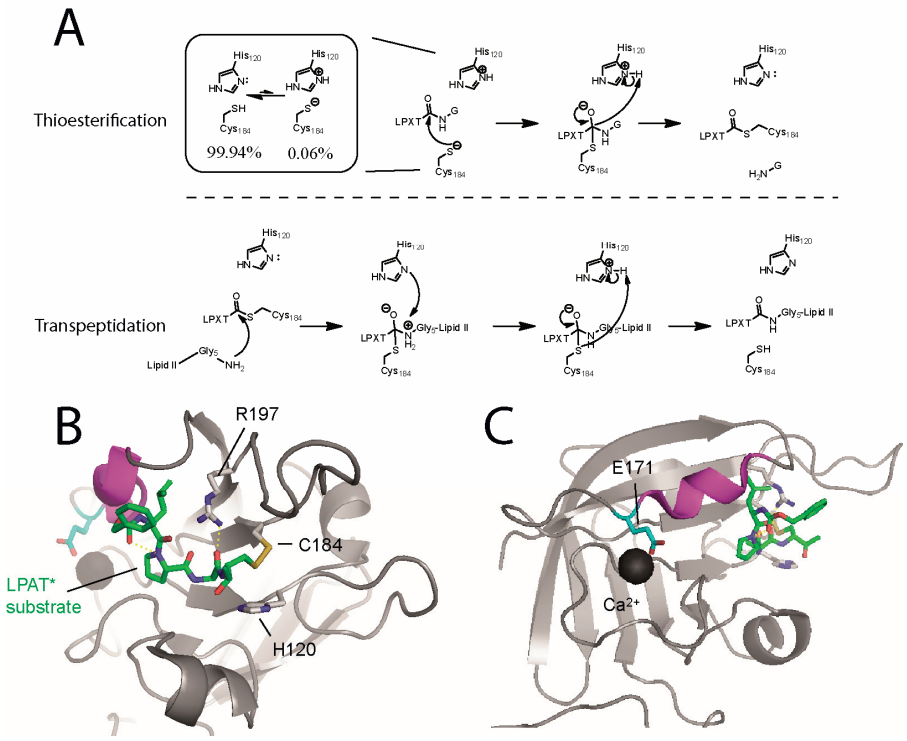


Figure 10. (A) Proposed reversed protonation catalytic mechanism for SrtA. The nucleophilic Cys184 thiolate attacks the carbonyl of the scissile T-G bond, resulting in the formation of a tetrahedral intermediate. His120 facilitates the collapse of the tetrahedral intermediate and formation of the acyl-enzyme by protonating the substrate leaving group. Figure adapted from Ref. 109. (B) and (C) NMR structure of the acyl-enzyme intermediate with a LPAT* substrate (green) (PDB ID 2KID). Ca^{2+} stabilizes the β6/β7 loop by orienting Glu171, inducing a short helix (magenta) providing contacts with the substrate.

Importantly, the expression of sortases, MSCRAMMs and other virulence factors is regulated in a growth-phase dependent manner by *agr* and *sar* loci[110]. Both surface proteins and sortases are predominantly expressed during exponential growth, while secreted proteins are upregulated during the stationary phase. This sequential expression may have clinical importance, as different stages of staphylococcal infection seem to require different panels of virulence determinants.

1.6.5 Sortase A inhibitors

SrtA has been targeted in several antibacterial drug discovery programs[111], and a number of inhibitors have been reported, although none has reached the clinic yet.

Several natural products have shown inhibitory activities against SrtA, the best in the micromolar range (Table 2, Figure 11A). Examples include aaptamines, topseptins and pigments such as curcumin and morin[112-114]. These natural products have shown effects

on adhesion and/or biofilm formation in *S. aureus* and other species when applied in the *in vitro* IC_{50} range. However, it is unclear whether these effects were due to the specific inhibition of SrtA or caused by off-target effects, since these compounds have also been shown to inhibit unrelated proteases or to induce cytotoxicity at the used concentrations[115-117].

The screening of chemical libraries has also led to the identification of small molecule inhibitors of SrtA. An example is DMMA (Figure 11B) with an IC_{50} = 9.1 μM[118]. This compound is the only SrtA inhibitor that has been evaluated in mice[119]. Mice treated with DMMA showed an increased survival rate and lower *S. aureus* titers in kidneys and joints. However, a similar effect was also observed in mice infected by *S. aureus* lacking SrtA, and therefore the reduced virulence could not be specifically attributed to SrtA inhibition. Inhibition of SrtB (for which DMMA has an IC_{50} = 34 μM) and other processes may have played a role. The most potent competitive inhibitor reported is a pyrazolethione compound[111] (Figure 11B). It showed sub-micromolar *in vitro* activities against SrtA but was reported to be unstable and its effects on *S. aureus* have not yet been assessed.

		IC_{50} *in vitro*[a]	**Assay on *S. aureus*[b]**	**Type of inhibition**	**MIC[c]**	**Ref.**
Cys-protease inhibitors	p-Hydroxy mercuribenzoic	120 μM	Seb anchoring	Covalent	74 μM	120
	Methanethiosulfonate	N.D.	Seb anchoring	Covalent	N.D.**	120
Natural products	Isoaaptamine	16.2 μM	Newman adhesion to Fn	N.D.	220 μM	112
	Bromodeoxytopsentin [bis(indole)alkaloid]	48 μM	Newman adhesion to Fn	N.D.	250 μM	113
	Curcumin	37.5 μM	Newman adhesion to Fn	N.D.	> 500 μM	114
	Morin (flavonoid)	37.4 μM	Fg clumping	N.D.	> 300 μM	121
	β-sitosterol-3-O-glucopyranoside	30 μM	Newman adhesion to Fn	N.D.	> 300 μM	122
	Berberine chloride	23 μM	N.D.	N.D.	270 μM	123
Chemical libraries	DMMA	9.2 μM	Newman adhesion to Fn	Competitive	> 650 μM	118,124
	Pyrazolethione compound	0.3 μM	N.D.	Competitive	> 500 μM	111
	Phenyl vinyl sulfone	740 μM	Newman adhesion to Fn	Covalent	6 mM	125
	AAEK2	15 μM	N.D.	Covalent	N.D.**	126
Peptidic	Peptidyl diazo/chloro-methanes	N.D.	N.D.	Covalent	N.D.	127
	Phosphinic-peptidomimetic	10 mM	N.D.	N.D.	N.D.	128

Table 2. Summary of SrtA inhibitors. (a) IC_{50} values reported *in vitro* with recombinant SrtA expressed in *E. coli*. (b) Assays performed to evaluate SrtA inhibition in *S. aureus*. (c) Minimum inhibitory concentration. N.D: not determined. **: antimicrobial activity was observed but MIC was not reported.

In order to develop more specific inhibitors, various groups tried to design inhibitors based on the substrate. For example, substrate-derived affinity labels based on the

recognition motif coupled to a chemical warhead (LPAT-CHN_2 and LPAT-CH_2Cl) showed selective labeling of SrtA in crude cell lysates prepared from transfected *E. coli* and wildtype *S. aureus*. These inhibitors had low K_i values of 0.2 μM but inactivation rates were relatively poor (in the order of 10^{-2} min^{-1}). In a different approach, phosphinic peptidomimetics (Figure 11C) were designed to mimic the transition state of the acyl-enzyme intermediate. The resulting compounds proved useful for the characterization of the catalytic mechanism[128], but they had weak affinities in the millimolar range, limiting their application as therapeutics. In conclusion, there is still a need for specific and potent inhibitors of SrtA to evaluate it as an antibacterial target.

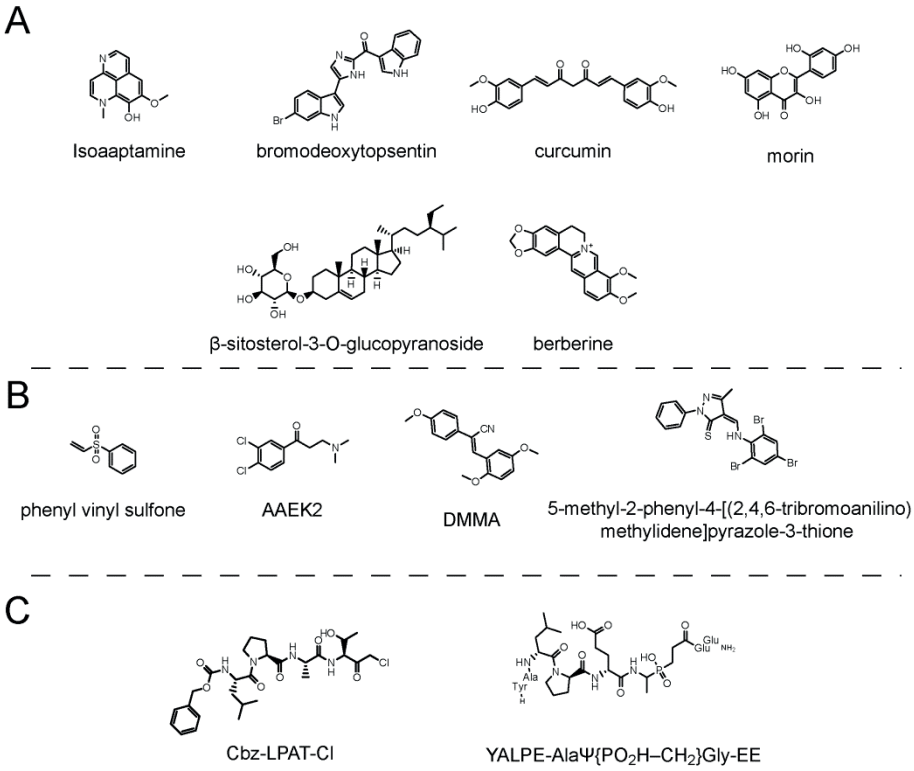


Figure 11. SrtA inhibitors. (A) Natural products. (B) Small molecule inhibitors from library screening and rational design. (C) Substrate-based peptidic inhibitors.

Chapter 2

Phage display libraries of differently sized bicyclic peptides

This chapter is based on a publication by Rentero Rebollo I, Angelini A, Heinis C. Phage display libraries of differently sized bicyclic peptides. **Med. Chem. Commun.** 2012, 4: 145-150 - Reproduced by permission of The Royal Society of Chemistry.

2.1 Introduction

Cyclic peptides are an enticing class of molecules for the development of therapeutics. They can act as protein domain mimics, reaching in some cases antibody-like affinity and specificity, while having favorable properties of small molecules such as high stability, access to chemical synthesis and good tissue penetration[4,129]. Phage display has emerged as a powerful tool for the high-throughput screening of polypeptides including cyclic peptides[130]. Since it relies on ribosomal synthesis, peptides are expressed as linear polymers and cyclization can be achieved by flanking the randomized segment with a pair of cysteine residues, which form a disulfide bridge upon oxidation. Recently, a new strategy has been developed to generate phage-encoded libraries of bicyclic peptides[43]. In brief, random peptides containing three cysteine residues are displayed on phage and reacted with the trivalent thiol-reactive compound tris(bromomethyl)benzene (TBMB) (Figure 12A). TBMB reacts efficiently and selectively with the cysteine side chains under mild conditions and yields bicyclic peptide structures[56]. Affinity selections against plasma kallikrein, cathepsin G and, more recently, urokinase-type plasminogen activator (uPA) yielded highly selective inhibitors in the nanomolar to sub-nanomolar range[43,44,62]. These bicyclic inhibitors had higher binding affinities than monocyclic ones previously generated to the same targets.

Loop length diversity is already an acknowledged component for the engineering of molecular recognition surfaces in protein scaffolds or cyclic peptides. Length plays an important role in the diversity of complementarity-determining regions of natural antibodies, and it is the primary determining factor for their conformations. Insertion of length diversity in synthetic antibody libraries allowed higher binding affinities to be

reached[131]. Similar results were obtained when the length of the binding loops was varied in other protein scaffolds such as the 10th type III domain of human fibronectin (Fn3)[132]. Furthermore, conformational diversity, given by loop length variability, was sufficient to compensate for restricted chemical diversity when screening Fn3 libraries having loops consisting of only Tyr and Ser residues[133]. Studies on cyclic peptide libraries suggested that the tightest binders are more likely to be identified by screening multiple libraries with variable loop length since targets often have preferences for specific peptide constraints[134,135].

So far, only macrocycle libraries having two rings of equal size have been screened by phage display[43,44,62]. Most affinity selections have been performed with a library having two equal rings, each containing six random amino acids (abbreviated as 6×6 bicyclic peptides)[43,44]. Phage panning against the serine protease plasma kallikrein yielded potent inhibitors[43]. Comparison of the peptide sequences revealed three consensus regions present in the first or second peptide loop. To affinity mature these inhibitors, three semi-randomized libraries were created, each having one of the three consensus sequences in one loop and six random amino acids in the other loop. Several improved inhibitors with K_is as low as 2 nM were obtained, wherein all improved clones were derived from only one of the three libraries. In selections against uPA, two classes of bicyclic peptides were isolated, the major one with a consensus sequence in the first peptide ring, and the minor one with a consensus sequence around the middle cysteine[44]. The best inhibitor (UK18) showed a K_i of 53 nM. In this case, affinity maturation attempts with libraries based on either of the consensus sequences could not improve the potency beyond the one of the best inhibitor. These results suggested that some peptide leads are more suited for the affinity maturation than others and that it is of advantage to have several consensus sequences as starting points.

In this work, we aimed at generating libraries of bicyclic peptides with different combinations of ring sizes in order to find more potent inhibitors and/or a larger diversity of binding motifs that could be used as constant regions in affinity maturation libraries. In a recent study, the size of bicyclic peptides was reduced to modulate their specificity. These bicyclic peptides having slightly smaller rings (5 instead of 6 randomized amino acids per ring, termed 5×5 peptides) bound tightly to the serine protease plasma kallikrein, inhibiting the human and murine orthologs but not any human paralogous proteases[62]. Herein, we generated bicyclic peptide phage libraries with combinations of differently sized macrocyclic rings. Specifically, we cloned 14 phage peptide libraries of the format Cys-$(Xaa)_m$-Cys-$(Xaa)_n$-Cys, wherein the number 'm' and 'n' of random amino acids between the cysteine residues was 3, 4, 5 or 6 (Figure 12). The libraries were subjected to affinity selections either in groups or all together against the cancer-associated protease uPA.

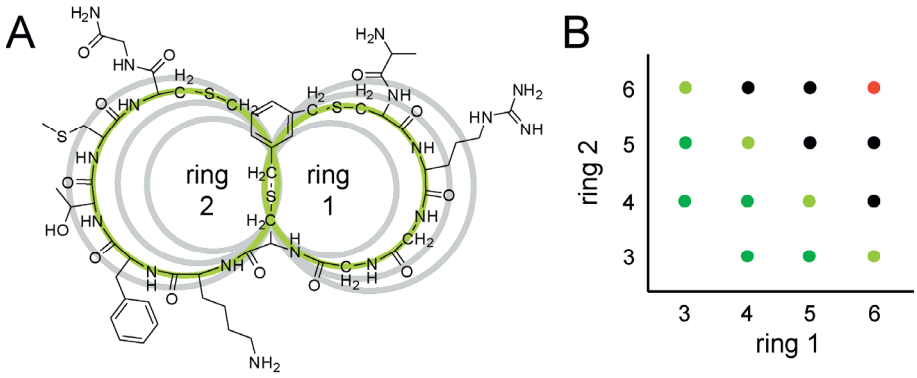


Figure 12. Bicyclic peptide phage libraries with different ring sizes. (A) Chemical structure of a representative bicyclic peptide (UK368) with green rings of 3 and 4 amino acids. For size comparison, the peptide rings with 3, 4, 5 and 6 variable amino acids are shown as grey circles. (B) Overview of the libraries. Indicated on the axes is the number of variable amino acids in the two rings of the bicyclic peptides. Libraries that were cloned in this work are indicated with light green (library A), green (library B) and black dots (library C). The 6×6 library indicated with a red dot was cloned previously.

2.2 Results and discussion

2.2.1 Phage-encoded bicyclic peptide libraries with variable ring sizes

The 14 bicyclic peptide phage libraries were cloned using degenerate primers that allow all 20 amino acids in the randomized positions. The libraries are termed according to the number of random amino acids per peptide ring. For example, the library 3×4 contains bicyclic peptides with 3 and 4 random amino acids in the first and second ring, respectively (Figure 12A). The library 6×6 already existed and was not newly cloned (Figure 12B, indicated as a red dot). Libraries with similar numbers of randomized amino acid positions were pooled as follows: bicyclic peptides of the format 3×4, 4×3, 4×4, 3×5 and 5×3 formed the library A (7 and 8 random amino acids; Figure 12B, light green dots), those of the format 3×6, 6×3, 4×5 and 5×4 formed the library B (9 random amino acids; Figure 12B, green dots) and those of the format 4×6, 6×4, 5×5, 5×6, 6×5 formed the library C (10 and 11 random amino acids; Figure 12B, black dots). The diversities of the libraries were quantified by counting the number of bacterial colonies formed after transformation and were relatively small (between 10^7 and 5×10^8 cfu). Sequencing clones from libraries A, B and C showed that peptides with all possible combinations of ring sizes were represented (Table 3).

2.2.2 Phage selections of bicyclic peptides against a serine protease

Phage selections were performed against the human serine protease uPA. The three libraries A, B and C modified with TBMB[56] were individually subjected to affinity selec-

tions. In parallel, a fourth selection was performed mixing the three libraries with the previously generated 6×6 library (size: 4×10^9 clones)[43], in order to allow competition among all the bicyclic peptide formats. Biotinylated uPA was immobilized on magnetic streptavidin beads in the first round and on magnetic neutravidin beads in the second round to prevent the enrichment of streptavidin or neutravidin binders. Negative controls in which the libraries were panned against streptavidin- (first round) or neutravidin- (second round) coated beads allowed the quantification of phage that bound specifically to uPA. Already in the first round of selection, the number of phage isolated against uPA was 10-fold higher compared to the negative control selection (Figure 13A). The enrichment over the negative control rose to 10^4-fold in the second round. Similar enrichments were observed for all libraries, suggesting that all contained a large portion of uPA-specific bicyclic peptide binders.

Library	Loop length	Before selection	2nd round	3rd round
A	3×4	7	3	12
	4×3	2	1	0
	4×4	8	4	7
	3×5	10	1	1
	5×3	6	3	1
B	3×6	4	1	0
	6×3	1	2	0
	4×5	15	12	8
	5×4	8	1	0
C	4×6	10	1	5
	6×4	4	4	6
	5×5	13	2	0
	5×6	3	1	0
	6×5	7	3	1

Table 3. Representation of bicyclic peptide formats. Indicated are the numbers of different clones with the respective format found in the library before and after two or three rounds of phage selection

2.2.3 Consensus sequences of bicyclic peptides isolated after two rounds of selection

Sequencing of 48 peptides isolated from the three libraries revealed a total of six different consensus sequences (Figure 14). One of the consensus sequences ($^S/_T$AR) was found in all the three libraries and in a total of 22 different peptides. This sequence appeared in either the first or second peptide ring, and was found in peptide rings with 4, 5 and 6 but not 3 amino acids length (Figure 14). The conformational constraint imposed in a loop with 3 amino acids might prevent the tri-peptide from binding to uPA. For the other five consensus sequences, bicyclic peptides within each group had in most cases the same or similar ring sizes. In peptides isolated from library A, two consensus sequences were identified, the first one being present in the second 4-amino acid ring of

3×4 and 4×4 peptides ($^{K}/_{R}$FSX; 'X' represents any amino acid). The second consensus sequence was found exclusively in bicyclic peptides having the 5×3 format ($^{S}/_{T}{}^{I}/_{L}$RCPSFC). In peptides isolated from library B, a consensus sequence was found in the second ring of 4×5 bicyclic peptides (TK$^{Y}/_{F}$T$^{L}/_{M}$). In selections with library C, a consensus sequence was found in bicyclic peptides of the 6×4 format (CNXYYSXC$^{S}/_{T}$). In the fourth experiment in which all libraries as well as the library 6×6 were mixed and subjected to selections, none of the bicyclic peptide formats was enriched over the others. However, the more ubiquitous $^{S}/_{T}$AR motif, which can be accommodated in different ring sizes, predominated in this selection.

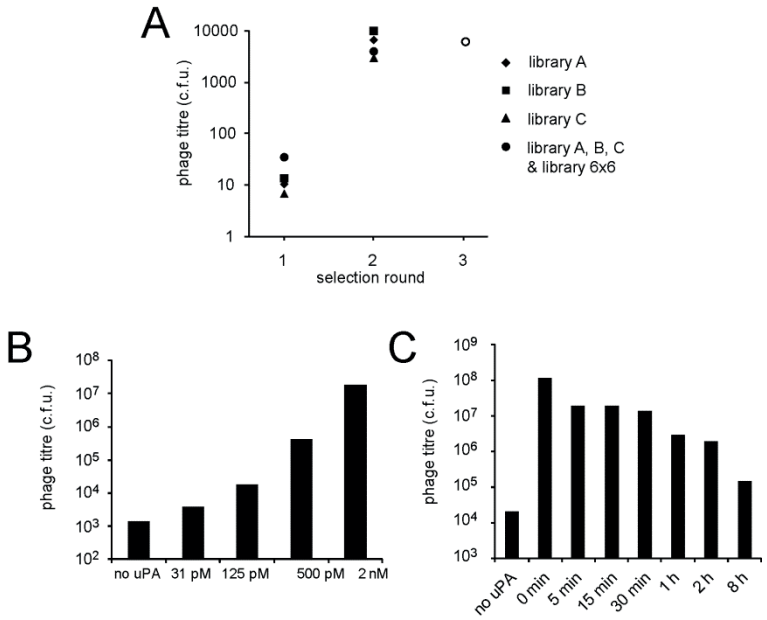


Figure 13. Number of phage clones isolated from the various libraries in the 3 rounds of selection. (A) Enrichment factors found for the different selections (number of phage captured in the presence of uPA divided by the number of phage isolated in the absence of uPA). The enrichment factors of 10^3 to 10^4 obtained for all libraries indicate that uPA-specific peptides were isolated. In the third round, phage isolated in all the four individual second round selections were pooled, amplified and together subjected to a third round of panning towards uPA. (B) Phage selections performed with different concentrations of uPA (third round of panning). The numbers of isolated infective phage particles are indicated. (C) Phage selections with competitive ligand (UK18) (third round of panning). After binding of phage to immobilized uPA, the competitive inhibitor UK18 (K_i = 53 nM)[44] was added for different incubation time periods (indicated on the abscissa in the graph) to prevent re-binding of weak ligands. The numbers of isolated infective phage particles are indicated.

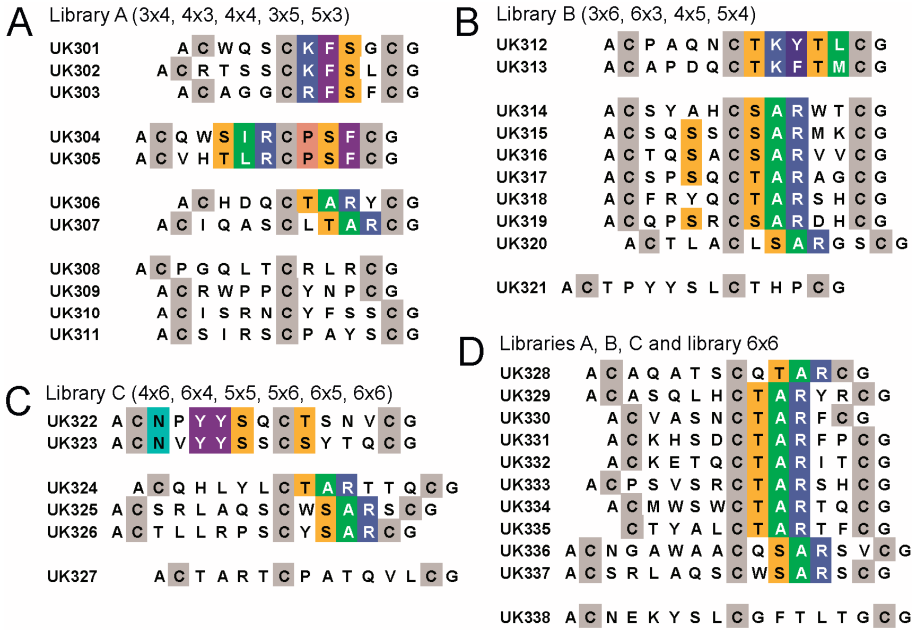


Figure 14. Bicyclic peptides isolated after two rounds of affinity selection. Libraries with similar numbers of randomized amino acids were pooled and subjected to selections with uPA. (A) Library A (7 or 8 randomized amino acids). (B) Library B (9 randomized amino acids). (C) Library C (10 or 11 randomized amino acids). (D) Library A, B and C as well as library 6×6. Sequence similarities are highlighted with colors.

2.2.4 Third round of panning applying more stringent selection conditions

Phage clones isolated after two rounds of panning in the four independent selections were mixed together and subjected to a third round (Figure 13A). To isolate only the tightest-binding peptides, we applied more stringent panning conditions. In one set of experiments, phage was incubated with lower concentrations of biotinylated uPA (ranging from 2 nM to 31 pM) before capturing them on streptavidin-coated beads. At lower concentrations, the number of captured phage was gradually reduced (Figure 13B). In the second set of experiments, the bicyclic peptide UK18, which binds to the active site of uPA with nanomolar affinity, was used to compete off weak binders. As anticipated, fewer phage remained on the beads at longer incubation times (Figure 13C). This latter experiment indicated that a large portion of the bicyclic peptides isolated in the second round of selection was binding to the active site of uPA.

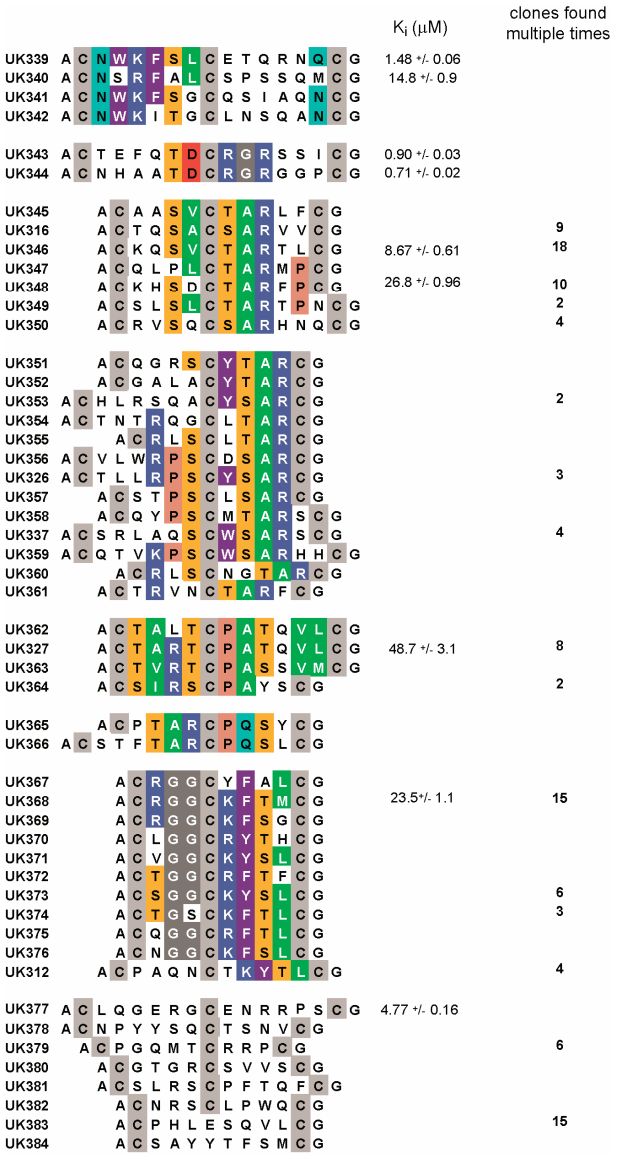


Figure 15. Bicyclic peptides isolated after three rounds of affinity selection. Sequence similarities are highlighted in color. The inhibitory activities (K_is) of several TBMB-cyclized peptides are indicated (average values of at least three measurements).

2.2.5 Consensus sequences of bicyclic peptides isolated after three rounds of selection

Sequencing of 149 clones isolated in the third selection round showed a smaller but still high sequence diversity (50 different peptides) and significantly stronger consensus sequences (Figure 15). Two of the consensus sequences derived from the 6×6 library were exactly the same as found in previous selections against uPA (Figure 15, top two consensus sequences). Several peptides contained again the tri-peptide motif ($^{S}/_{T}$AR)

that was already found in different bicyclic peptide formats after the second round of panning. This consensus sequence was flanked by specific amino acids that likely increase the binding affinity. A strong consensus covering 9 amino acid positions ($CT^{A}/_{V}RTCPA^{T}/_{S}XV^{L}/_{M}C$) and containing the TAR motif was preferentially found in the 4×6 bicyclic peptide format with one exception, a 4×4 bicyclic peptide (UK364). Another consensus sequence also containing the TAR motif covered 6 amino acid positions (TARCPQS) and was found in bicyclic peptides of the 4×4 and 6×4 formats. A strong consensus sequence not containing the TAR motif was found in 10 different peptides all having the 3×4 format ($CXGGC^{K}/_{R}{}^{F}/_{Y}{}^{T}/_{S}{}^{L}/_{M}C$).

Synthetic bicyclic peptides of several clones were chemically synthesized. All of them inhibited uPA but none was more potent than the previously isolated UK18 (Figure 15). The most potent inhibitors isolated here, UK343 and UK344, were derived from the 6×6 library and showed K_is of 0.90 and 0.71 μM, respectively. It is likely that the 6×6 library had yielded the best binders because of the larger number of randomized amino acid positions and/or the larger number of different clones in the library (4×10^9 clones versus 10^7 to 5×10^8 clones in the libraries A, B and C). Although more potent inhibitors could not be isolated from the naïve libraries, the numerous binding motifs identified provide more starting points for future affinity maturation attempts. Of particular interest are the bicyclic peptide formats with large rings that appear, based on the consensus sequences, not to be optimized in all of the amino acid positions.

2.3 Conclusion

Phage panning experiments with bicyclic peptides having different ring sizes yielded more diverse consensus sequences than previously found in selections with a 6×6 bicyclic peptide phage library. The bicyclic peptide inhibitors with unrelated consensus sequences are presumably interacting differently with the active site of uPA. Some of these peptides may bind in orientations that allow affinity maturation of non-conserved regions while others do not. Having available multiple leads isolated from such bicyclic peptide libraries with variable ring sizes could therefore be a great asset for the generation of high affinity binders. It is likely that other multicyclic peptide structures evolved by phage display such as cysteine knots[136] or other disulfide-constrained miniproteins[137-139] would similarly benefit from variation of the peptide ring size.

2.4 Experimental procedures

2.4.1 Library generation

All primers used for library cloning, as well as the vector 21tet(5), are described in the Supplementary Information found in APPENDIX I. Phage libraries were created by

inserting DNA sequences encoding the semi-random peptide sequences (Ala-Cys-$(Xaa)_m$-Cys-$(Xaa)_n$-Cys, *m* and *n* ranged from 3 to 6 amino acids), the linker Gly-Gly-Ser-Gly, and the disulfide-free domains D1 and D2 into the phage vector 21tet(5). The insert was step-wise created in two consecutive PCR reactions. First, the genes of D1 and D2 were PCR amplified with the two primers prepcr and sfi2notfo using the vector fdg3p0ss2[38] as a template. Second, the DNA encoding the random peptides was appended in a PCR reaction using primers of the type 5′-TATGCGGCCCAGCCGGCCATGGCAGCATGC$(NNK)_m$TGC$(NNK)_n$TGTGGCGGTTCTGGCGCTG-3′ (m, *n* = 3, 4, 5 or 6) and the primer sfi2notfo. The PCR products were digested with *Sfi*I (underlined in primers) and ligated into *Sfi*I-digested vector 21tet(5). For each set of libraries (A, B and C), 23 µg and 7 µg of *Sfi*I-digested vector and PCR products respectively were ligated and electroporated into *E. coli* TG1 cells. An equal amount of each insert was added into the ligation reaction to a total of 7 µg. After electroporation, cells were incubated for 1 hour in 2YT at 37 °C and plated on large (20 cm diameter) chloramphenicol (30 µg ml^{-1}) 2YT plates. Colonies were scraped off the plates with 2YT media, supplemented with 10% v/v glycerol and stored at −80 °C.

2.4.2 Phage selections

Phage were produced and the peptides modified with TBMB as described previously[4]. Chemically modified phage were dissolved in 3 ml washing buffer (10 mM Tris–Cl, pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$, 1 mM $CaCl_2$), and blocked by addition of 1.5 ml washing buffer containing 3% w/v BSA and 0.3% v/v Tween 20, for 30 min. In parallel, 5 µg of biotinylated uPA was immobilized by incubation with 40 µl magnetic streptavidin beads (Dynabeads M-280 Streptavidin, Invitrogen, Switzerland). The beads with uPA as well as 40 µl magnetic streptavidin beads without uPA were blocked for 30 minutes in separate tubes with blocking buffer (washing buffer containing 1% w/v BSA and 0.1% v/v Tween 20). The 4.5 ml blocked phage were distributed equally to the two tubes and incubated for 30 minutes on a rotating wheel. The beads were washed eight times with washing buffer containing 0.1% v/v Tween 20 and twice with washing buffer. The bound phage were eluted by incubation with 100 µl of 50 mM glycine pH 2.2 for 5 min. Eluted phage were transferred to 50 µl of 1 M Tris–Cl, pH 8.0 for neutralization, and incubated with 30 ml of exponentially growing TG1 cells (OD_{600} = 0.4) for 90 min at 37 °C. The cells were pelleted by centrifugation, dissolved in 1 ml 2YT, spread on 2YT plates containing chloramphenicol (30 µg ml^{-1}) and incubated at 37 °C overnight. The colonies were scraped off the plates with 2YT media, supplemented with 10% v/v glycerol and stored at −80 °C. In the second round of panning, uPA was immobilized on magnetic NeutrAvidin beads. The beads were prepared by coating tosyl-activated beads (Dynabeads M-280 Tosyl-activated, Invitrogen) with NeutrAvidin (Pierce, Rockford, IL, USA) following the manual of the manufacturer. In the third selection round, phage isolated in the second round of the four independent selections were produced separately, mixed at equal numbers, blocked as described above and panned together against uPA. In the in-solution capture procedure, biotinylated uPA (10 nM) was blocked, and incu-

bated at final concentrations ranging from 31 pM to 2 nM with 2 ml of blocked phage. After incubation for 45 min at room temperature with rotation, the phage/uPA–biotin complexes were captured by incubation for 7 min with 40 µl blocked streptavidin beads. The washes and the elution were performed as described above. In the competitive capture procedure, 120 ng biotinylated uPA was immobilized on 30 µl streptavidin beads, washed twice and incubated in 300 µl blocking buffer for 30 min. They were subsequently added to 2 ml blocked phage and incubated at room temperature with rotation. Bicyclic peptide UK18 was added at different time points to a concentration of 9 µM. The phage were washed, eluted and propagated as described above.

2.4.3 Chemical synthesis of bicyclic peptides

Peptides were synthesized by standard solid-phase peptide synthesis using Fmoc-protected amino acids and Rink amide AM resin (see Experimental procedures in APPENDIX I). Peptides were eluted under reducing conditions and partially purified by precipitation. Crude peptide (0.5 mM) was reacted with TBMB (1 mM) in 80% aqueous buffer (20 mM NH_4HCO_3, 5 mM EDTA, pH 8.0) and 20% CH_3CN for 1 h at 30 °C. The product was purified by reversed-phase chromatography on a C18 column (XBridge BEH300 Prep 5 µm, Waters, Milford, MA, USA) using a linear gradient with a mobile phase composed of eluant A (99.9% v/v H_2O, 0.1% v/v TFA) and eluant B (94.9% v/v CH_3CN, 5% v/v H_2O and 0.1% v/v TFA) and a flow rate of 20 ml min^{-1}. Pure bicyclic peptides were lyophilized and dissolved in H_2O. The molecular mass was confirmed by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry.

2.4.4 Determination of inhibitory activity

Human uPA at a final concentration of 4 nM was incubated with different concentrations of bicyclic peptides and 100 µM fluorogenic substrate (Z-Gly-Gly-Arg-AMC, Bachem, Bubendorf, Switzerland) and residual activity measured at 25 °C for 30 minutes. The reactions were performed in volumes of 150 µl and a buffer containing 10 mM Tris–Cl, pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$, 1 mM $CaCl_2$, 0.1% w/v BSA, 0.01% v/v Triton-X100 and 5% v/v DMSO at 25 °C. Fluorescence intensity was measured with a Spectramax Gemini fluorescence plate reader (excitation 355 nm, emission 480 nm, Molecular Devices). The reactions were performed in triplicate. The final K_i was calculated using the Cheng–Prusoff equation[140], wherein the kinetic constant K_m for the hydrolysis of Z-Gly-Gly-Arg-AMC by human uPA was determined by the standard Michaelis–Menten equation as previously described[44].

Chapter 3

Identification of target-binding peptide motifs by high-throughput sequencing of phage-selected peptides

This chapter is based on a publication by Rentero Rebollo I, Sabisz M, Baeriswyl V, Heinis C. Identification of target-binding peptide motifs by high-throughput sequencing of phage-selected peptides. **Nucleic Acids Res.** 2014. Reproduced with permission of Oxford University Press.

3.1 Introduction

Phage display of peptides is widely used for the development of peptide ligands and for epitope mapping[141,142]. The procedure involves 2-4 iterative rounds of affinity selection and phage amplification followed by an optional ELISA-based screen and sequencing of several dozens of positive clones. A panel of peptides is synthesized and their binding to the protein target or biological activity tested. An important step in the phage selection of peptides is the comparison of sequences and the identification of consensus motifs. Consensus sequences can provide valuable information about the binding site of peptides. Peptides sharing the same consensus motif likely bind to the same surface region of the target protein and form similar molecular interactions. Multiple different consensus sequences indicate that peptides bind with different interaction modes to the same or different surface regions. Selections with peptide libraries often yield only one consensus sequence or at maximum a few different ones. In many phage selections with peptide libraries, no consensus sequences are reported at all. If isolated ligands are to be used as leads in drug development, multiple consensus sequences are desired as parallel development of several peptide leads increases the success rate of the development program. For example, peptides of one consensus sequence might share unfavorable properties such as poor solubility or low proteolytic stability, hindering their further development.

The sequences of phage-selected peptides are typically obtained by Sanger sequencing. Our laboratory, for example, is routinely sequencing a half or a whole 96-well plate of clones isolated after 2-3 rounds of phage panning. Sequence similarities among peptides

are identified by manual comparison of the sequences, and highlighted by coloring amino acids of aligned peptides or by representation as so-called logos. In recent years, high-throughput sequencing (HTS) methods have been applied for the analysis of ligands isolated from DNA-encoded chemical libraries[74,143], or antibodies[76,144], protein domains[77,145,146] and peptides[75,79,82,147-150] isolated from phage display or mRNA display libraries. Most of the sequencing work was done using Roche's 454 sequencing technology (earlier work)[74,144,145,147], the Illumina platform[75-77,79,82,143,146,148-150], or an Ion Torrent sequencer[79]. The vast sequence data gave valuable information about diversity and abundance of isolated clones, as well as allowed monitoring of these parameters during the different iterative rounds of selection and amplification. In selections of peptide ligands, the sequencing data was analyzed primarily by ranking the peptides according to their abundance, and the most frequent peptides were characterized. 't Hoen *et al.*[82] and Olson *et al.*[77] showed that peptide ligands can be identified in a single round of selection. In order to distinguish functional clones over background, Olsen *et al.* subjected each clone in > 1000 copies to the selection (input) and identified potential binding sequences from the 10 most abundant peptides. Herein, we proposed to analyze high-throughput sequencing data of phage-selected peptides not only based on abundance, but also based on sequence homology. We expected that sequencing and comparison of ten-thousands of peptides could allow a finer discrimination of consensus sequence sub-groups. Extensive sequence homology information could provide information about binding interactions and the importance of specific residues for the binding.

Powerful tools to compare extensive sequence data and to identify multiple different consensus sequences within large datasets of sequences have been developed. The algorithms of MEME (Multiple Em for Motif Elicitation)[150], MUSI (MUltiple Specificity Identifier)[151] and Gibbs Cluster[152] can process large numbers of sequences and group them in clusters of similar peptides. The three tools unfortunately do not provide information about frequencies and nucleotide sequences in the analysis result. Derda and co-workers developed MatLab-based software for the analysis of phage-selected peptides sequenced by the Illumina platform[75]. The tool tailored for the commercial Ph.D.™ -12 Phage Display Library (New England Biolabs) provides information about sequence abundance and DNA sequences but it does not include a function for automated identification of sequence homologies.

In this work, we conceived a strategy to identify target-binding peptide motifs by high-throughput sequencing and sequence comparison. We developed a procedure and software for vast data processing, sequence quality filtering and homology finding. We applied it to bicyclic peptides that were isolated against five different protein targets. The tools allowed identification of numerous sub-families of consensus sequences. We show that target-binding peptide motifs can be identified even after only one round of affinity selection.

3.2 Results

3.2.1 Phage selection and high-throughput sequencing

Bicyclic peptide phage libraries were generated by displaying linear peptides of the format $X_lCX_mCX_nCX_o$ (C = cysteine; X = any amino acid; l, m, n, o = number of random amino acids) on filamentous phage and subsequent chemical cyclization of the peptides with tris-(bromomethyl)benzene (TBMB)[153]. The libraries were panned against the five targets sortase A from *S. aureus* (SrtA), human urokinase-type plasminogen activator (uPA), activated human coagulation factor XII (FXIIa), human plasma kallikrein (PK) and streptavidin (SA). Bicyclic peptide libraries were previously screened against four of these targets (uPA, FXIIa, PK, SA) and had yielded binders with micromolar to picomolar dissociation constants[58,62,63,154]. In these previous selections, consensus sequences were identified by sequencing around 100-300 clones per target after 2-3 iterative selection rounds. Against the bacterial target SrtA of *S. aureus*, bicyclic peptides were not developed so far. Isolated peptides were analyzed after a single round of phage selection instead of after 2-3 iterative rounds, as usually done. A single round of selection minimizes out-competition of weaker binders by stronger ones. In this way, a maximal number of target-binding peptide motifs was expected to be identified. Conversely, a single round of selection bore the risk that binders were not sufficiently enriched over non-binders, making the identification of consensus sequences more difficult. Bicyclic peptide libraries with different format (ring sizes of 3 to 6 amino acids), different complexity (10^7 to 4×10^9 different clones) and different representation of individual phage clones (ranging from 2 to 1000 copies per clone) were applied as shown in Table 4. In some selections, phage clones were represented in high copy numbers to facilitate enrichment of individual clones over non-specifically selected 'background' peptides. This was expected to facilitate analysis of data and identification of consensus sequences. After one round of phage selection, phage DNA was sequenced on an Ion PGM™ Sequencer instrument using an Ion 316™ Chip, yielding a maximum of 5×10^6 reads per chip. This number was exceeding by far the number of phage isolated in the phage selections, ranging from 4×10^2 to 3×10^4 (Table 4). It even allowed sequencing phage from multiple selections on a single chip. DNA of selected phage was isolated from bacterial cells and amplified by PCR using suitable primers as shown in Figure 16A and APPENDIX II-Table S2. A 6-letter barcode was included in the forward primers right after the adaptor sequence to allow multiplexing of up to ten different phage selections on a single chip. Samples run on an Ion 316™ Chip yielded more than a million reads and thus more than 100,000 sequences per phage selection (Table 4).

Target	Library[a]	Library diversity[b]	Phage input[c]	Phage output[c]	Total reads	Total sequences[d]	Different sequences[e]	% top 200[f]
SrtA	Library A	5×10^8	3×10^{10}	8×10^3	1.8×10^5	5.1×10^4	2.8×10^3	26%
SrtA	Library B	1×10^7	2×10^{10}	1×10^4	2.4×10^5	6.8×10^4	1.4×10^3	75%
uPA	Library B	1×10^7	9×10^9	3×10^4	3.4×10^5	1.1×10^5	3.1×10^3	56%
FXIIa	4×4	7×10^8	5×10^{10}	1×10^4	4.1×10^5	1.7×10^5	7.9×10^3	15%
PK	3×3, 4×4	1×10^9	2×10^9	2×10^3	1.8×10^5	7.5×10^4	1.4×10^3	40%
SA	3×3, 4×4	1×10^9	2×10^9	4×10^2	1.1×10^5	6.1×10^4	3.4×10^2	84%

Table 4. Summary of the protein targets and peptide phage display libraries. [a]Libraries are named according to Ref. 153. Library A contains 3×4, 4×3, 4×4, 3×5, 5×3 peptides, library B contains 3×6, 6×3, 4×5, 5×4 peptides. [b]Number of transformants. [c]Transducing units (t.u.). [d]Total number of sequences after quality filter. [e]Estimated number of different sequences. [f]Percentage of the population corresponding to the top 200 clones.

3.2.2 Data processing and analysis

We developed MatLab software for the processing and analysis of sequence data as outlined in the flow diagram shown in Figure 16B. Several of the applied procedures such as sorting of sequences, quality filtering, abundance ranking and translation were based on scripts developed by Derda and co-workers[75]. In a first step, sequences provided by the Ion Torrent sequencer in fastq format[155] were distributed into different files according to their barcode to separate peptides from different phage selections. For barcodes having a single base mutation, deletion or insertion, a correction function was developed but it proved to rescue only a small fraction of peptides and was not further used (described in APPENDIX II-Supplementary Data and Figure S4). In a second step, low-quality sequences were removed from the dataset, and identical sequences sorted by their abundance. In the same step, the DNA sequences were translated into amino acid sequences. The software allows specifying the start and end of the region to be evaluated, so that it can be applied to any peptide library, regardless of the length of the random sequence and the flanking residues. In a third step, peptides were optionally sorted according to their format (i.e. number of cysteines and number of residues between them) or based on inter-dataset comparisons (e.g. peptides isolated in two independent phage selections). In a fourth step, peptides were pairwise compared to find consensus sequences and to identify target-binding motifs. In a fifth and last step, identified peptide motifs were used to search the entire dataset for more related sequences. In all processes, information about peptide abundance and nucleotide sequence is displayed. All descriptions of the scripts are available in the APPENDIX II. The scripts can be used for the analysis of any files in fastq format and thus also for data sequenced with other technology platforms such as Illumina.

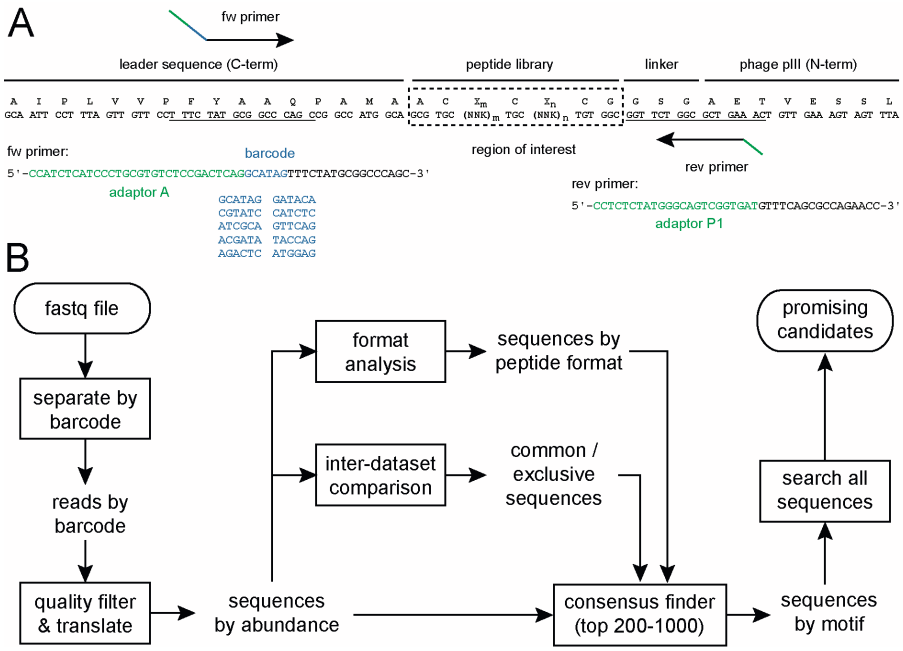


Figure 16. High-throughput sequencing and sequence analysis strategy. (A) Primer design for Ion Torrent sequencing of bicyclic peptide libraries (Library A and Library B). (B) Procedure for the analysis of sequencing data applying MatLab scripts. First, reads are separated into several files according to their barcode. Second, low-quality sequences are removed from the dataset, and remaining sequences are translated and sorted by abundance. At this point, two optional steps are performed: distribution of the isolated peptide sequences based on the format (e.g. peptide ring size in the case of bicyclic peptides) and comparison of two different datasets. Then, the sequences of the most abundant peptides (e.g. top 200) are compared and clustered in consensus groups or subfamilies of consensus groups, allowing the identification of specific motifs. Finally, the entire pool of sequences is searched for other less abundant sequences sharing such motifs in order to identify promising candidates.

3.2.3 Reducing bias by optimizing quality parameters

A critical step in the data processing is the filtering of sequencing data based on quality criteria. Ion Torrent is prone to over-calling or under-calling the length of homopolymeric regions, leading to insertion/deletion (indel) errors[156]. The confidence of each sequenced nucleotide (Q-value) is provided in the fastq file with a single-character Phred-based quality score[155], assigned by the PGM base-caller. This information can be used to remove sequences containing low confidence basecalls prior to sequence analysis. Application of too strict quality filters, however, can lead to a bias against homopolymer-containing sequences. Different quality filter stringencies were tested and an optimal one chosen. A filter allowing a maximum of three nucleotides having a quality score below 'Q18' was found to be optimal for all datasets. The importance of optimal quality filtering is illustrated in Figure 17 in which different quality filters were applied to peptides isolated from the 4×4 library against PK (Figure 17A): a "permissive" filter, in

which reads containing 3 nucleotides below quality score 'Q18' were discarded, and a "restrictive" filter, where only 1 nucleotide below quality score 'Q20' was allowed. Although the difference in the total number of reads passing each filter was minimal (Figure 17C), certain peptide sequences were completely lost when using the restrictive filter (Figure 17A). DNA of such peptides contained a tetra-thymine homopolymer (TGT-TTT; encoding Cys-Phe) as well as a penta-thymine homopolymer (TGT-TTT-TCT; encoding Cys-Phe-Ser) (Figure 17B). We observed similar biases in all datasets. In order to reduce the bias against sequences containing long homopolymers, the less strict quality filter with a maximum of three nucleotides under the quality score 'Q18' was applied.

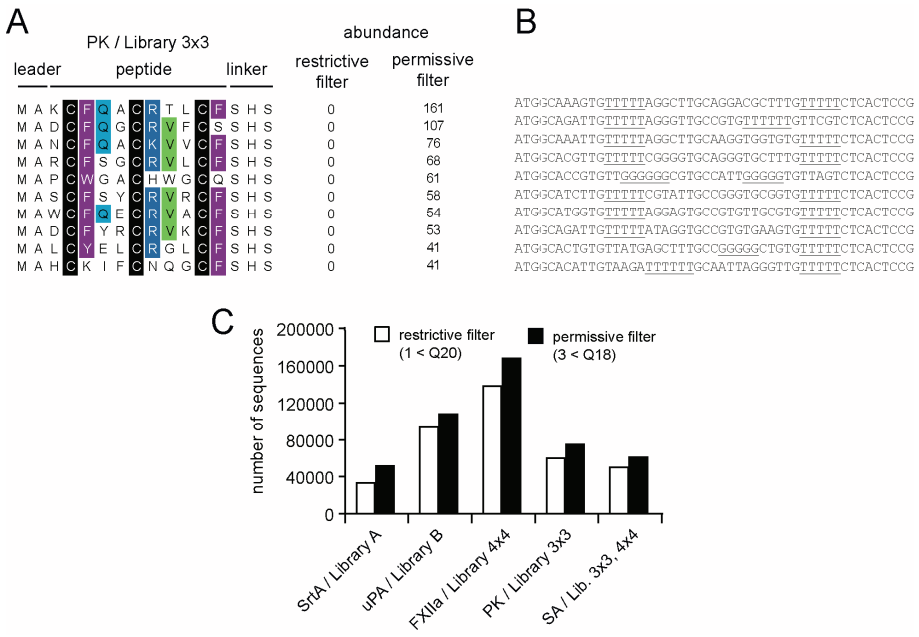


Figure 17. Application of an optimal sequencing quality filter. Comparison between permissive (a maximum of 3 bases with quality value lower than Q18 are allowed) and restrictive (a maximum 1 base with quality value lower than Q20 is allowed) filtering parameters. (A) Example of peptides rescued by applying a less restrictive quality filter to the selection against PK. The rescued peptides with the highest abundance are indicated (top 10). (B) DNA sequence of rescued peptides. The homopolymers in the DNA sequences are underlined. (C) Effect on the number of reads passing the different filtering parameters.

3.2.4 Diversity of phage-selected peptides

The copy number of the most abundant peptides varied strongly among different selections. In some selections, the 200 most frequently found peptides represented more than 80% of the sequenced clones, while in other selections they formed a fraction of less than 20% (Figure 18A). We plotted the number of different peptide sequences against the number of analyzed reads to extrapolate the absolute number of different peptide

sequences found in each selection. We expected that the number of different sequences converges to a maximal value at larger numbers of analyzed peptides and fits to equation 1 where *a* is the total number of different peptide sequences in the dataset and *k* is a constant that depends on the abundance distribution of the sample. Equation 1 was found to be suitable for fitting simulated datasets containing (i) different numbers of peptides, and (ii) different peptide abundance distributions (APPENDIX II-Supplementary Data, Figure S5 and Figure S6).

$$\text{Equation (1)} \qquad y = a\left(1 - e^{-\frac{x}{k}}\right)$$

The number of different sequences increased linearly at larger numbers of sequences analyzed and did not converge to a maximal value, as well as did not fit to equation 1. The linear increase was due to sequencing errors, which were directly proportional to the number of sequences. Taking this phenomenon into account, we fitted the data to equation 2 where *a* and *k* are again the total number of different peptide sequences in the dataset and a constant that depends on the abundance distribution of the sample, respectively, and *b* is the average error rate of the population. Equation 2 was also verified with simulated datasets containing (i) different number of peptides, (ii) different abundance distributions, and (iii) different error rates (APPENDIX II-Supplementary data and Figure S7).

$$\text{Equation (2)} \qquad y = a\left(1 - e^{-\frac{x}{k}}\right) + bx$$

Data of all selections was fitting well to equation 2 (Figure 18B). The linear coefficient *b* was similar in datasets of all selections. Experimental datasets of this study contained a significant percentage of sequencing errors estimated to be between 2.8% and 5.1%. The number of different sequences calculated for the various selections ranged between 340 and 8,000 and was hence consistent with the number of isolated phage (Table 4). The different peptides isolated in selections could thus essentially all be identified by sequencing around 100,000 clones.

After one round of phage selection, we expected that propagation advantages of specific clones would not have a large impact on the selection results. To evaluate the extent of the propagation-related bias after one round of selection, phage of library A and B were produced and bacterial cells infected without affinity selection. The copy number of individual clones increased only marginally and the most abundant clones represented in both cases less than 0.02% of the population (APPENDIX II-Figure S8) and were not found after selection.

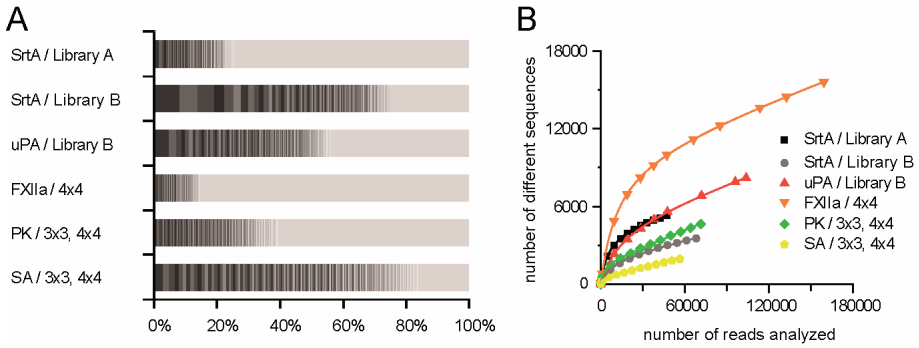


Figure 18. Diversity of peptides isolated after one round of phage selection. (A) The abundance of the 200 peptides that were most frequently found is indicated in percent of the whole population of sequenced clones (indicated in blocks colored in different grayscales). (B) Number of different sequences found when increasing numbers of reads were analyzed. Saturation plots were used for the calculation of the total number of different sequences.

3.2.5 Identification of target-binding peptide motifs

Based on MatLab built-in functions, we developed a script that groups peptides according to similarities. First, it calculates pair-wise distances among the peptides. It then constructs a phylogenetic tree using the distances calculated. Last, it clusters the peptides in suitable groups, with two optional parameters to fine-tune this grouping (see script description in APPENDIX II). This script allowed to efficiently identify target-specific binding motifs. The MatLab script generated well-arranged groups of around 3-20 peptides with high sequence similarity that can be analyzed and validated by eye.

Inspection of consensus groups revealed that some of them were not true consensuses but artifacts that resulted from sequencing errors as explained in the following. For highly abundant peptides, peptide variants with nearly the same sequence were found. These peptides occurred in small copy numbers and typically differed in only one base from the abundant clone (e.g. insertion, deletion or mutation). An example from a selection against SrtA using library A is shown in Figure 19A: the most abundant clone was present 4592 times and several clones with similar DNA sequences appeared in only a few copies (ranging from 8-26). It is likely that the low-copy sequences resulted from sequencing errors because peptides with such small sequence differences are unlikely represented in the library. For example, library A contains only a small fraction (around 10^8 different peptides) of the theoretically possible sequences (around 10^{12} sequence calculated from 8 positions encoded by NNK codons). To eliminate sequencing errors and prevent false identification of consensus sequences, we developed a MatLab script that finds sequences that differ at only one or two positions and corrects them to the sequence of the more abundant clone. Indeed, application of this script led to elimination of a significant fraction of the errors. Consensus sequence artifacts were no longer found.

Additionally, after this correction, the parameter *b* in equation 2 decreased below 1% (Figure 19B and APPENDIX II-Table S3).

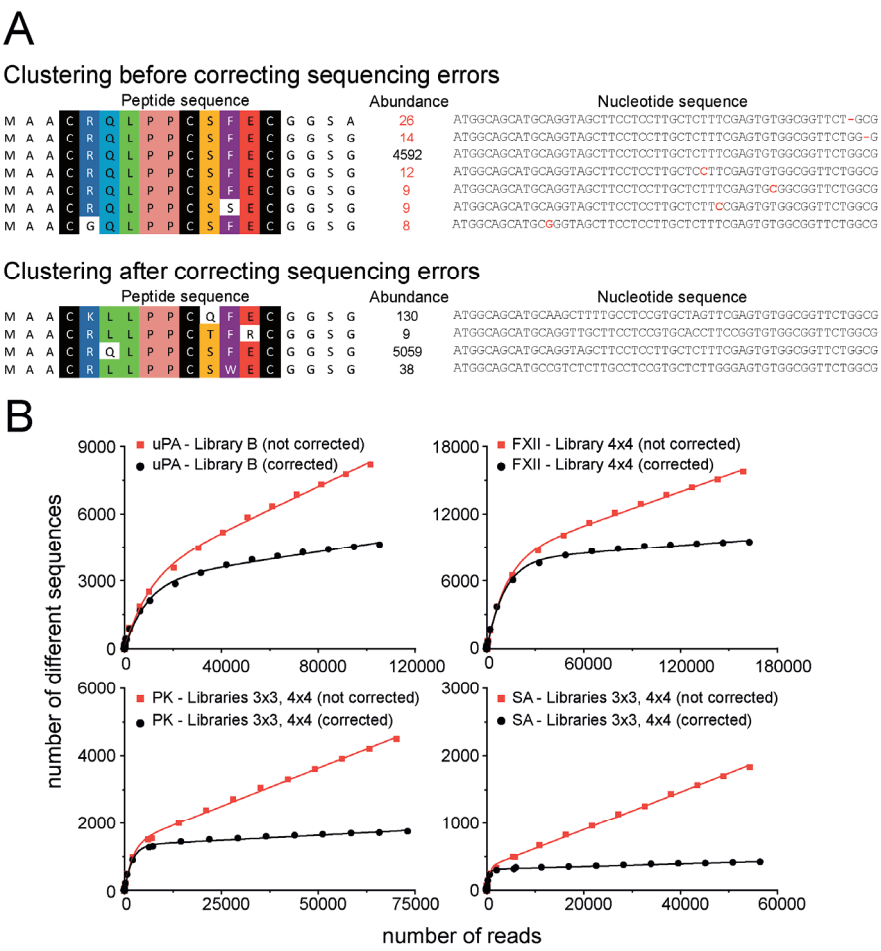


Figure 19. (A) Example for the identification of false consensus sequences due to sequencing errors. In a selection of Library A against SrtA, the most abundant sequence (present 4592 times) was clustered with sequences differing in only one nucleotide and being present at much lower frequency. These sequences likely resulted from sequencing errors. A MatLab script (fixingerrors.m script) was developed to eliminate these erroneous sequences. For the high-abundance sequence shown in the figure, 467 erroneous sequences were found (9%). For other high-abundance sequences, wrong sequences ranged between 0 and 48%. (B) Examples of saturation plots for different datasets before and after correcting sequencing errors (all datasets were obtained after one round of phage selection).

Consensus groups found after elimination of false sequences in the selections against all five protein targets are shown in Figure 20. As the required computational power increases quadratically with an increasing number of peptides, we compared only the top 200 abundant sequences from the different datasets. This was sufficient to identify consensus motifs in all selections. The analysis of larger numbers of sequences (up to 1000

sequences), did not lead to the identification of more target-binding motifs in this work (data not shown), but it may do if applied to other selections. In all phage selections performed, groups of peptides with high sequence similarities were found. Many of the groups formed by the MatLab script represented subfamilies of a few entirely different consensus sequences. We manually highlighted the sequence similarities in all consensus groups with color (Figure 20).

Consensus sequences shared by only a small number of peptides were identified too. For example, the SA-binding motif HPQ was shared by as little as 3 different peptides in the SrtA selection and was still identified by the software. These peptides were isolated because biotinylated SrtA was immobilized on SA in the phage selection. In the uPA selection, the minor motif '$^{K}/_{R}{}^{F}/_{Y}{}^{S}/_{T}$L' was shared by 9 different peptides. The peptides could be assigned even to two different consensus sequence subfamilies (Figure 20).

In all the selections, at least one or two target-binding peptide motifs could be found, namely 'LPP' for SrtA, '$^{T}/_{S}$AR' and '$^{K}/_{R}{}^{F}/_{Y}{}^{S}/_{T}$L ' for uPA, 'VxxKCL' for FXIIa, '$^{F}/_{Y}\backslash^{W}$ xxCRV' for PK and 'HPQ' for SA (Table 5). The number of different consensus sub-families was much larger; it was 15 for SrtA, 16 for uPA, 2 for FXIIa, 11 for PK, and 2 for SA. The motifs identified in selections with uPA, FXIIa, PK and SA were previously found by us or others after iterative rounds of phage selection and peptides with these motifs proved to be binders[58,62,63,154]. In contrast, most of the consensus sub-families had not been previously identified. The peptide motif 'LPP' found in selections against SrtA was not reported before; synthetic peptides with this motif bound to SrtA (results are described in Chapter 4). Searching the whole pool of sequenced peptides for the identified target-binding peptide motifs revealed many additional sequences that are potential ligands of interest for characterization. Some consensus sequences contained up to around 2000 different peptides (e.g. in the uPA selection). Other contained as little as 93 different peptide sequences (FXIIa selection). In some selections, peptides with binding motifs represented more than 50% of the total number of sequenced peptides (uPA selection) or as little as 1% (FXIIa selection).

Target	Library	Peptide motifs	Number of subfamilies	Different peptides with motifs in top 200	Different peptides with motifs in whole pool	% population containing a binding motif
SrtA	Library A	LPP	15	143 (72%)	1531 (41%)	47%
SrtA	Library B	LPP	14	164 (82%)	1253 (54%)	81%
uPA	Library B	$^{T}/_{S}$AR	14	165 (82%)	1943 (42%)	70%
		$^{K}/_{R}{}^{F}/_{Y}{}^{S}/_{T}$L	2	8 (4%)	44 (1%)	2.7%
FXIIa	4×4	RPCP	1	2 (1%)	23 (0.2%)	0.4%
		VXXKCL	1	5 (2%)	93 (1%)	1.2%
PK	3×3, 4×4	$^{F}/_{Y}\backslash^{W}$XXCRV	11	90 (45%)	758 (43%)	43%
SA	3×3, 4×4	HPQ	2	17 (8.5%)	37 (9%)	7.4%

Table 5. Target-binding peptide motifs (patterns of conserved residues) found after one round of selection.

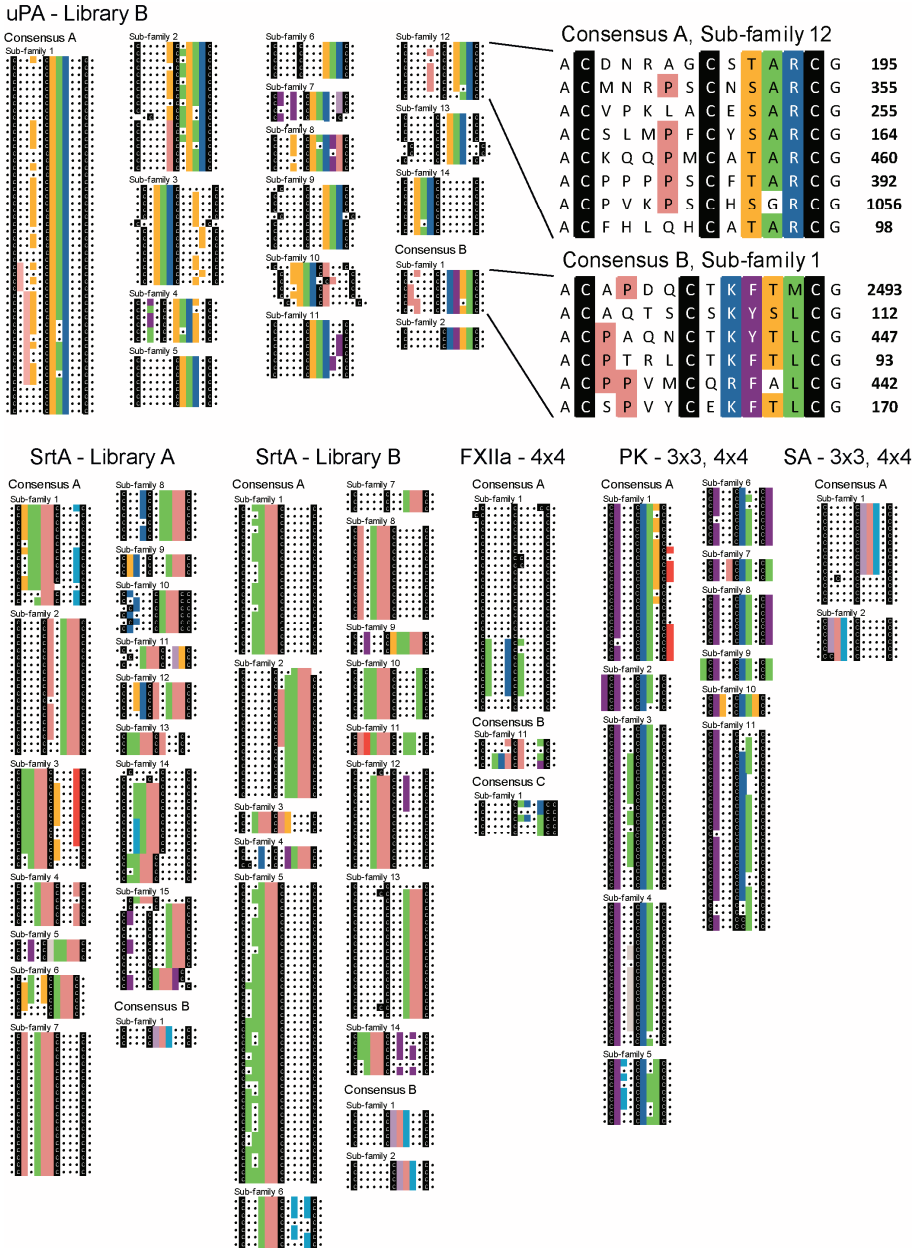


Figure 20. Identification of target-binding peptide motifs. The 200 most abundant peptides of each selection were computationally compared and clustered into groups of peptides that share a maximal sequence similarity. The raw data of the automated sequence comparison is included in the APPENDIX II. The sub-groups generated computationally were arranged manually to group those together that belong to the same consensus group. The cysteines are colored in black and regions in the peptides with sequence similarities were manually highlighted in color. Top: Consensus groups of peptides isolated in the selection with uPA. Peptide sequences of two of the sub-groups are enlarged and shown together with the abundance on the right side. Bottom: Consensus groups of peptides isolated against SrtA, FXIIa, PK and SA.

3.2.6 Peptide motif identification from inter-dataset comparisons

In phage panning experiments, many phage particles are isolated unspecifically (named background phage) along with the peptides that are selectively isolated through binding to a target. If the number of specifically isolated peptides is small compared to the unspecific ones, it is more difficult to identify specific target-binding sequences after just one round of selection. Additional rounds of phage selection may be needed. We hypothesized that, in such cases, a possible way to identify specific target-binding sequences in the presence of high background would be to perform two parallel selections and compare the sequences obtained. Identical peptides would be considered as target-specific peptides. We repeated a first round of selection against FXIIa and found that only six peptide sequences were common in both pools. Four of them corresponded to confirmed binding motifs that were previously found after three rounds of selection (Table 6)[58].

Abundance selection 1	Abundance selection 2	Peptide sequence	Peptides identified in previous phage selections*
307	12	ACDARPCPQTYCL	yes
40	110	QCVPLKCLWDRCE	yes
27	22	VCERQVCYLMSCW	no
12	36	TCLCKRCIKELCC	yes
11	16	YCVWDKCLWLMCE	no (but similar to consensus)
5	9	ACGMSICVLYGCN	no

Table 6. Peptides identified by inter-dataset comparison. (*) In previous phage selections, three iterative rounds of panning were performed and around 100 clones sequenced.

3.2.7 Formats of isolated peptides

The number of cysteines found in phage-selected peptides can indicate if they are forming linear, monocyclic or bicyclic peptide structures. We anticipated the isolation of peptides with three cysteines that are cyclized with TBMB and form bicyclic peptide structures. Occasionally, peptides with less or more than three cysteines are isolated from the applied phage peptide libraries. Previous work showed that peptides having a fourth cysteine residue in the randomized region are isolated as bicyclic peptides formed by two disulfide bridges[63]. Due to errors in the library generation, some peptides have two cysteines and are isolated as disulfide-linked monocyclic peptides. Availability of the vast sequence data allowed detection of small differences in the number of cysteines and preferences for one or the other format in the different selections. In selections performed with libraries containing peptides of different ring sizes (number of amino acids spacing the cysteines), we analyzed if one or the other format was preferentially isolated. Peptides with certain ring sizes were preferentially enriched in selections with some protein targets. In the selection of SrtA binders from library A, bicyclic peptides of the formats 3×5 and 5×3 were enriched over other formats (Figure 21A). Panning of li-

brary B against SrtA enriched bicyclic peptides of the format 5×4, while panning against uPA yielded more bicyclic peptides of the format 4×5 (Figure 21B). When the libraries 3×3 and 4×4 were mixed and panned against PK, 3×3 clones had a selective advantage, which was not the case when the same mix was panned against SA (Figure 21C).

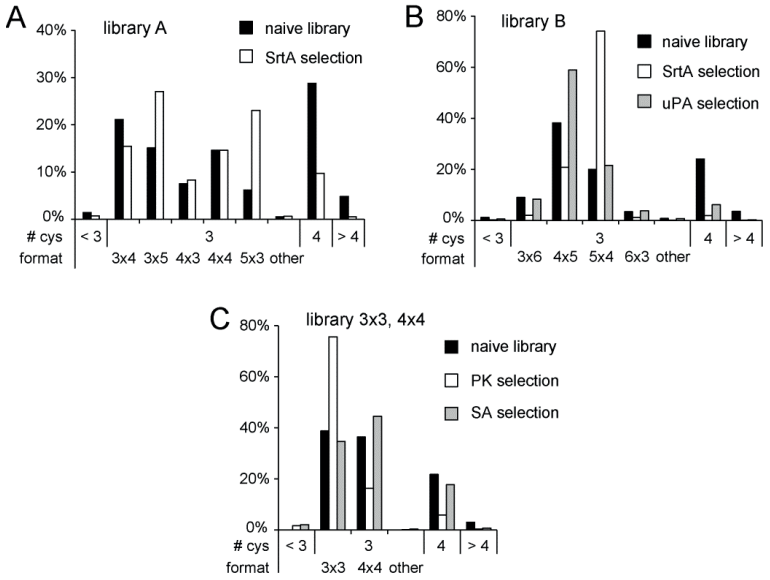


Figure 21. Statistical analysis of peptide formats using large sequence data. The percentages of peptides containing 0, 1, 2, 3, 4 or > 4 cysteine residues are indicated. For peptides containing 3 cysteines, the percentage of peptides with different formats are indicated. For example '3×4' means that the peptides contain 3 amino acids between Cys1 and 2, and 4 amino acids between Cys2 and 3. (A-C) Results of selection with different targets and libraries. 'Naive' means the peptides in the library before selection.

3.2.8 Iterative rounds of phage selection

We performed a second round of phage selection to study the population diversity (number of different sequences) and homogeneity (abundance distribution) over two rounds of selection. In particular, we were interested to learn (i) how many sequences with consensus motifs are lost in a second round of selection, and (ii) if new sequences with binding motifs appear. Phage isolated from library A and library B against SrtA in the first round were subjected to a second round of affinity selection against SrtA and isolated clones sequenced. The population underwent a progressive loss of diversity over iterative rounds of selection (Figure 22). The number of different sequences decreased from 2800 (round 1) to 800 (round 2) in the case of library A, and from 1400 to 170 in the case of library B. In the selection with library A, around half (47%) of the peptides isolated in round 1 contained the 'LPP' motif and thus were binders. In round 2, nearly all the peptides (98.4%) were binders. Around one third of the sequences with the binding motif 'LPP' found in round 1 were lost in round 2. Interestingly, 22% of the population of

the second round corresponded to sequences that were not found in the first round, indicating that the sampling of the first round was not complete and not all the diversity of the first round was sequenced. In the selection with library B, in the first round already 81% of the population of reads corresponded to binding sequences. After the second round, virtually all the population consisted in target-binding sequences (99.4%). A large fraction of the population after round 1 was also found in round 2 (71%), and new binding sequences found in the second round corresponded to less than 1% of the population.

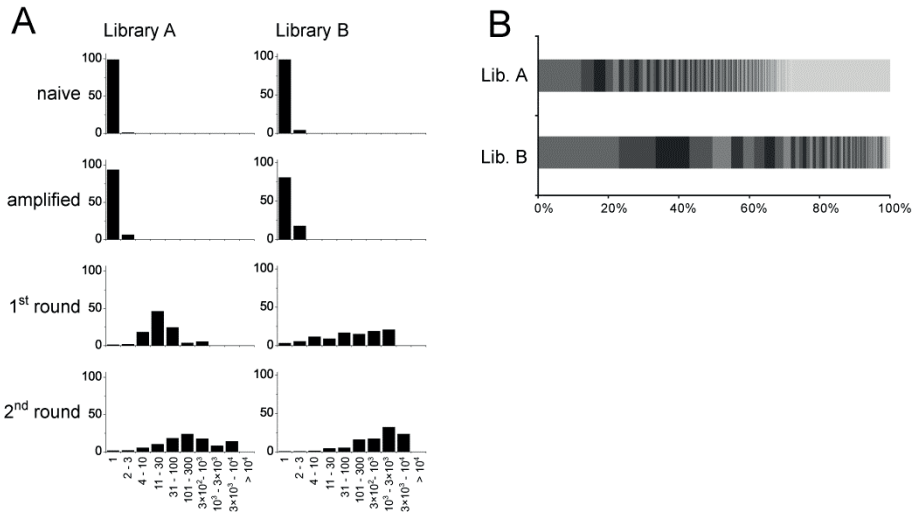


Figure 22. Dynamic change of the population over two rounds of selection. Results are shown for the selection of SrtA binders from libraries A and B. (A) Copy number of sequenced peptides. Indicated is the percentage of peptides that were identified at the indicated range. (B) Abundance distribution of the output of the second round. The most abundant 200 peptide sequences are separated in blocks.

3.3 Discussion

Sequencing of phage-selected peptides by high-throughput methods can offer a deep insight into the nature of selected peptides and the process of affinity panning and propagation. Pioneering studies in which phage-selected peptides were sequenced with high-throughput methods primarily used the data to study the peptide diversity and to identify highly abundant clones that are expected to bind with the highest affinities[75,79,82,147-150]. Herein, we proposed to use high-throughput sequencing data to identify target-binding motifs as well as to obtain a more detailed picture of consensus sequences. A limitation we encountered was the lack of broadly applicable and flexible open-access computational tools to compare and analyze the sequences of a large number of peptides. We therefore devised a procedure and developed software that processes high-throughput sequencing data and that can identify consensus sequences.

In our strategy, phage-selected peptides are first ranked by their abundance and then compared pairwise to align peptides with sequence similarities. The software reads sequence raw data from fastq files that are provided by most high-throughput sequencing platforms. The output of the tool consists on groups of 3-20 peptides sharing sequence similarities. Importantly, the software is keeping the information about the abundance and nucleotide sequence of each peptide sequence and displays this information in the analysis result. The software can deal with commercially available as well as self-tailored libraries. It includes functionalities for analysis of specific library formats such as disulfide-cyclized peptides or bicyclic peptides. Additional functions allow inter-dataset comparisons as well as searching for peptides containing specific sequence motifs.

While developing the analysis procedure and software, we learned that it is important to understand biases introduced by next-generation sequencing technologies. It is paramount to optimize quality filters to prevent introduction of biases. The main error source in Ion Torrent PGM sequencing is inaccurate flow-calls, which result in insertion/deletion (indel) errors, most frequently in homopolymeric regions[157,158]. Even correctly called homopolymeric regions are typically assigned less confidence[156]. Filters applied inappropriately could remove too many sequences and in this way introduce strong biases. We empirically identified an optimal quality filter which tolerates three bases with qualities below Q18. This filter gave the best result for all selections presented in this work and most likely is suitable for analysis of peptides isolated from any other type of combinatorial peptide library.

Sequencing errors were found to mislead standard algorithms that are used to identify sequence similarities and consensus sequences. We show that sequencing errors on highly abundant clones produce a series of erroneous variants, whose abundance is generally lower. The abundant clone together with a group of similar erroneous sequences were recognized by the software as a consensus group. We developed a procedure that eliminates sequencing errors from the dataset. False sequences are identified as such if they have identical nucleotide sequences except for one or two positions. Application of this filtering procedure eliminated the identification of false consensus sequences.

Our software was able to identify consensus sequences and sub-families of consensus sequences in datasets of all phage selections. Even consensus motifs that were shared by only a few peptides in the population could be identified. As we compared only the most abundant 200 peptides in each selection, some consensus motifs were most likely missed. More target-binding motifs may be identified if significantly more peptides are compared. The scripts were run on a standard personal computer within minutes. Thousands of sequences may be compared by using high performance computers. In the analyzed 200 sequences per selection, 1-3 consensus sequences were found that were further divided into many sub-families with slight consensus variations. This finding indicated that most proteins have only one or at most few regions where peptides can bind with sufficiently high affinity allowing their isolation. This is in contrast to antibodies that typically bind to more different epitopes.

An important parameter in the phage selection is the copy number of the peptides that are subjected to affinity selections. Only if a peptide is available in the library in a sufficiently large copy number, it can be isolated and sequenced in multiple copies and appears as an 'enriched' peptide. In some of the selections performed in this work, the average copy number of the peptides was rather low and the isolated peptides diverse. The identification of target-binding peptide motifs was thus difficult. For example in the selections against PK and SA, the average copy number was 2. Consensus sequences could in these cases only be identified because many peptides were sharing the same motif.

In selections with more challenging targets such as FXIIa, it was difficult to identify target-binding motifs. Only 1% of all peptides isolated against FXIIa contained FXIIa-binding motifs (some of the motifs were known from previous work). Most of the 99% remaining sequences are most likely peptides that were isolated through non-specific interactions. Our software could nevertheless identify two consensus sequences. We also investigated the possibility of reliably identifying specific-binding sequences by performing in parallel independent selections. We reasoned that this approach could allow the identification of specific target-binding ligands from noisy datasets and for the identification of parasitic sequences. By comparing the output of two selections performed in parallel against FXIIa (one selection round), we indeed could differentiate specific target-binding clusters from background clusters. Inter-dataset comparison may also be applied to identify peptides that bind to the streptavidin magnetic beads rather than to the protein target.

Our work confirmed that peptide ligands can be efficiently identified in a single round of phage selection if isolated clones are analyzed by high-throughput sequencing. In contrast to previous work that identified peptides ligands based on their abundance, we show that extensive comparison of sequences can identify additional attractive ligand candidates. Phage selection of peptide ligands in a single instead of multiple rounds has also the advantage that propagation-related bias is reduced to a minimum[75,159,160]. This could be particularly important when genetically engineered phage systems containing unnatural amino acids are used[161]. Finally, a single round of phage panning may also facilitate the application of phage display by scientists that have no prior experience with this technique. Readily prepared libraries could simply be pipetted to a target and captured phage sequenced. Phage amplification and purification would not be necessary, and equipment for bacteria culture and phage handling would not be required.

In summary, we have developed a strategy and software to compare large numbers of phage-selected peptides that were sequenced by high-throughput methods. With this strategy, we were able to identify rare target-binding peptide motifs, as well as to define more precisely consensus sequences and sub-groups of consensus sequences. This information is valuable to choose peptide leads for drug development and it facilitates identification of epitopes.

3.4 Experimental procedures

3.4.1 Phage selection

Libraries A, B, 3×3 and 4×4 were previously described[62,154]. In these libraries, peptides are displayed on around five copies of the phage coat protein pIII. Libraries A and B contain peptides of the format ACX_mCX_nCG (C = cysteine, X = any amino acid). In library A, the combinations of 'm' and 'n' are 3/4, 4/3, 4/4, 3/5 and 5/3; in library B they are 3/6, 6/3, 4/5 and 5/4. Library 3×3 contains peptides of the format XCX_3CX_3CX. Library 4×4 contains peptides of the format XCX_4CX_4CX. Random positions are coded by NNK codons. Phage production, reaction of cysteines with chemical linker to generate bicyclic peptides on phage, and phage panning against the different targets were performed as described before[58,62,154]. The vector for *S. aureus* sortase A expression pHTT14[162] was kindly provided by Prof. O. Schneewind (University of Chicago, IL, US). Sortase A was expressed in *E. coli* (amino acid 26-206, polyhistidine tag at N-terminus) and purified by nickel affinity chromatography followed by size exclusion chromatography. Human urokinase-type plasminogen activator N322Q was expressed in mammalian cells, activated and purified as described before[61]. Human coagulation factor XIIa (β-form) and human plasma kallikrein were purchased from Molecular Innovations (Novi, MI, USA). The proteins were biotinylated and immobilized on streptavidin magnetic beads (Dynabeads M-280, Life Technologies, Carlsbad, CA, USA). For streptavidin selections, the commercial streptavidin magnetic beads were readily used.

3.4.2 Sample preparation for high throughput sequencing

Phage vector was extracted from TG1 *E. coli* bacteria that were stored as glycerol stocks after infection with phage isolated after one round of selection. The DNA was isolated with a commercial plasmid purification kit (NucleoSpin Plasmid; Macherey-Nagel, Düren, Germany). 100 ng phage vector DNA was amplified by PCR using primers containing adapter sequences and barcodes (primer sequences are provided in APPENDIX II-Table S2). The PCR reaction in a volume of 50 µL contained final concentrations of 250 µM dNTP, 500 nM primer, 1 unit Taq polymerase, and standard buffer (75 mM Tris-HCl, 20 mM $(NH_4)_2SO_4$, 2mM $MgCl_2$, 0.01% Tween 20). 25 PCR cycles (30 sec. 95 °C, 30 sec. 55 °C, 30 sec. 72 °C) were performed but resulted in formation of DNA heteroduplexes and only about 30% of all sequences could be read. Reduction of the number of PCR cycles to 13 solved this problem. PCR products were purified from a 2.5% agarose gel (UltraPure agarose, Invitrogen, Carlsbad, CA, US) using a commercial agarose gel purification kit (NucleoSpin Gel and PCR Clean-up; Macherey-Nagel). The concentration of DNA was determined using a High Sensitivity DNA Assay Kit (Agilent, Santa Clara, CA, US), following the manufacturer's protocol. Ion Torrent sequencing was performed by the Lausanne Center of Genomic Technologies (University of Lausanne, Switzerland) or the Centre for Research in Agricultural Genomics (Barcelona, Spain) on a Ion Personal Genome Machine (PGM™) Sequencer. The procedure involved ligating the DNA frag-

ments onto Ion Sphere Particles (ISPs), amplifying them by emulsion PCR, enriching the templated ISPs, loading onto an Ion Torrent 316™ chip, and sequencing.

3.4.3 Analysis

MatLab scripts were developed for the analysis of high-throughput sequencing data (all descriptions can be found in the APPENDIX II). A first script, *Step1.m*, sorts the reads according to the specified barcodes and distributes them to separate files. Reads with mutations, insertions or deletions in barcodes were discarded unless specified. A second script, *Step2.m*, removes low quality reads, translates the sequences, sorts them by abundance, and optionally corrects sequencing errors. Reads having more than three bases with quality score lower than Q18 were not considered, unless specified otherwise. Sequences differing in one or two bases from an abundant sequence were corrected as the small differences likely origin from sequencing errors. MatLab scripts *LoopLengths.m*, *Clustering.m*, *FindSeq.m* and *CommonSeq.m* were used for the comparison and analysis of peptide sequences. Script *LoopLengths.m* separates the sequences into different files according to the number of cysteine residues and the number of amino acids between them. Script *Clustering.m* compares a chosen number of sequences, groups them into families that share high sequence similarity, and optionally generates sequence logos for each group. Script *FindSeq.m* searches the dataset for all peptide sequences containing a specified motif. Script *CommonSeq.m* compares up to three different datasets and distributes common and exclusive sequences in different files.

Chapter 4

Development of selective peptide macrocycle inhibitors of *S. aureus* sortase A

4.1 Introduction

The development of antibiotic resistance among life-threatening human pathogens has prompted the exploration of new alternative targets beyond those exploited by conventional antibiotics. A group of proteins that are considered as interesting novel targets are sortases. They are membrane-bound transpeptidases that catalyze the transfer and covalent immobilization of surface proteins to the cell wall in gram-positive bacteria. In *S. aureus*, one of the most relevant pathogens due to the existence of strains resistant to virtually all antibiotics in the clinic, two sortases have been described: sortase A (SrtA) and sortase B (SrtB)[100,101], which recognize two different motifs and therefore anchor different surface proteins. A number of important virulence factors such as protein A, clumping factors and fibronectin-binding proteins are anchored to the cell wall by SrtA. They enable adhesion and infection of host cells and tissues, evasion from the immune system and biofilm formation[163-165]. SrtA knockouts show reduced adhesion to matrix proteins and reduced pathogenicity in animal models for *S. aureus* infections[102-104,166], as well as for infections caused by other microorganisms[167,168]. Consequently, great interest has arisen in the development of SrtA inhibitors for therapy[164,169].

A range of small molecules have been reported to inhibit SrtA in cleaving fluorescence-based peptide substrates, including natural products, such as berberine chloride (IC_{50} = 23.3 μM)[123] and curcumin (IC_{50} = 37.5 μM)[114], and small molecules from chemical libraries and rational design, such as the diarylacrylonitrile-derived compound DMMA (IC_{50} = 9.1 μM)[124] and phenyl-vinyl sulfone (IC_{50} = 700 μM)[125]. Most of the compounds were found to inhibit adhesion of *S. aureus* to matrix proteins, and this activity was attributed to SrtA inhibition[114,124,125]. The therapeutic potential of one of the compounds, DMMA, has been evaluated in mice. *S. aureus* infections in mice treated with DMMA showed reduced virulence leading to higher survival rates[124]. Besides small molecules, a

peptide-based ligand of SrtA isolated from a lantipeptide library by mRNA display also showed good affinity (1.3 μM), but did not inhibit SrtA[170].

In this study, we aimed at developing potent and selective SrtA inhibitors based on peptide macrocycles for further evaluation of the target and for a potential use as leads in antibiotic drug development. We envisioned developing inhibitors based on bicyclic peptides that contain two macrocyclic rings for interaction with the target. Bicyclic peptide ligands with nanomolar or even picomolar affinity have recently been developed in our laboratory to a range of protein targets using a phage display-based approach[43,44,58,62]. In this work, we identified a series of SrtA bicyclic peptide inhibitors in the low micromolar range and applied them to evaluate the SrtA target of *S. aureus*.

4.2 Results and discussion

4.2.1 Phage selection of bicyclic peptide SrtA ligands

Three combinatorial peptide libraries of the form Ala-Cys-$(Xaa)_m$-Cys-$(Xaa)_n$-Cys-Gly (m, n = number of random amino acids) were displayed on phage and cyclized by reacting the cysteines with tris-(bromomethyl)benzene (TBMB) as previously described[153,154]. Library A contained bicyclic peptides with 7-8 random residues (m×n = 3×4, 4×3, 3×5, 5×3, 4×4), library B contained peptides with 9 random residues (4×5, 5×4, 3×6, 6×3), and library 6×6 contained bicyclic peptides with 12 random residues (6×6). After two rounds of selection, we analyzed the selected peptides both by Sanger sequencing of 30-35 clones (Figure 23A) and by high-throughput sequencing (HTS) (Figure 23B). In all cases, the most abundant clones identified by HTS (representing more than 4% of the population) were also identified by clone-picking. The 20 most abundant clones from each library are shown in Figure1B and additional clones in the APPENDIX III. Nearly all peptides isolated from libraries A and B contained the amino acid motif "LPP" (98% and 99% of the peptides, respectively). The motif was present at the C-terminal position of rings containing five amino acids (…C**LPP**C…), and appeared more often in the first one of the two rings. Peptides isolated from library 6×6 converged to two consensus motifs: around one third of the peptides (37%) contained a fourth cysteine and the consensus sequence "AC$XX^K/_R$XVCC$^L/_V$X$X^D/_E$XXCG" and half of the peptides (50%) contained the "LPP" motif in either of the two rings (Figure 23). Previous studies in our laboratory showed that peptides having four cysteine residues are likely isolated in the form of bicyclic peptides with two disulfide bridges[63]. Peptides with unmodified cysteines can oxidize during phage preparation or affinity selection. Although the fraction of unmodified peptide is typically low[171], peptides with two disulfide bridges can be enriched if they bind particularly well to the target protein. Analysis of all the sequences obtained by HTS did not lead to the identification of other motifs.

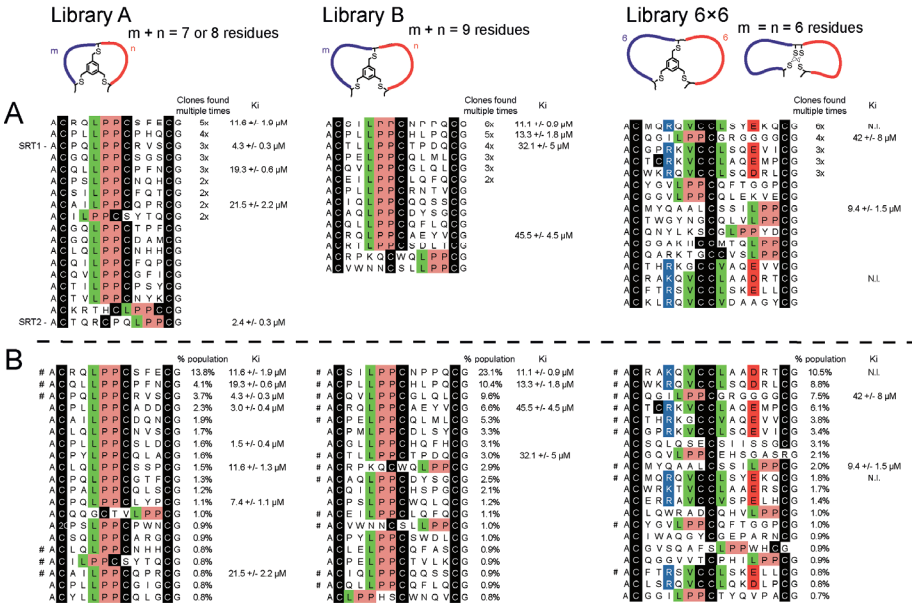


Figure 23. Phage selections against SrtA. Three bicyclic peptide libraries were panned against *S. aureus* SrtA. Selected peptides were identified after two rounds of selection, and their inhibitory activities were evaluated. K_is (averages of at least 3 independent measurements and standard deviations) for SrtA are indicated, N.I.: no inhibition was observed at the highest concentration tested (300 μM). (A) Sanger sequencing of 30-35 clones. Peptides found multiple times are indicated. (B) High-throughput sequencing using Ion Torrent PGM. The 20 most abundant sequences for each library are indicated, with their frequency (% of the population). Peptides found by clone-picking are indicated with a hash symbol (#).

4.2.2 Bicyclic peptide inhibitors of SrtA

We synthesized 16 bicyclic peptides and tested the inhibition of SrtA with a fluorescence-quenched substrate Dabcyl-LPETG-Edans. All LPP-containing peptides that were tested blocked SrtA with K_i values in the micromolar range (Figure 23). The "LPP" motif is similar to the sorting sequence LPXTG recognized by SrtA in its natural substrates and might bind in a similar manner as the substrate[172]. Conversely, no inhibitory activity was observed for peptides containing two disulfide bridges that were isolated from the 6×6 library. The best inhibitors, with potencies in the single-digit micromolar range, contained the extended motif "$P^{Q}/_{L}$LPP" in either the first or the second ring (Figure 23). Analysis of the HTS data showed that proline was indeed the most frequent amino acid in the first position of the 5 amino acid loop (Figure 24A). At the second position, leucine, valine and glutamine were the most frequent amino acids. In the SrtA inhibitors with the "$P^{Q}/_{L}$LPP" motif in the first ring, the second ring did not converge to a specific sequence (Figure 24A). In contrast, in the group of peptides with the "$P^{Q}/_{L}$LPP" motif in the second ring, some amino acids in the first ring appeared to be preferred. In order to evaluate whether the non-conserved loop was contributing to the binding, we performed an alanine scan with two peptides SRT1 (AC**PQLPP**CRVSCG, K_i = 4.3 μM) and SRT2

(AC<u>T</u>QR<u>C</u>**PQLPP**<u>C</u>G, K_i = 2.4 μM) (Figure 24C). Mutation of amino acids in the second ring of SRT1 did not affect the K_i, suggesting that the second ring was not interacting with the target. Mutation of amino acids in the first ring of SRT2 reduced the inhibitory activity substantially. Mutation of Gln4 and Arg5 to Ala reduced the activity 7 and 22-fold, respectively (Figure 24C). The preference for an arginine in position 5 was also visible from the HTS results analysis. The additional sequence requirements in amino acids positions 4 and 5 might have been the reason why fewer peptides were selected with the LPP motif in the C-terminal loop. We further characterized the binding affinity of these two prototypic peptides, SRT1 and SRT2, by fluorescence polarization. The peptides labeled at the N-terminal amino groups with fluorescein-NHS bound SrtA with K_d values of 8.8 μM and 1.5 μM, respectively (Figure 24D). These binding affinities were in line with the inhibitory constants K_i of the two peptides.

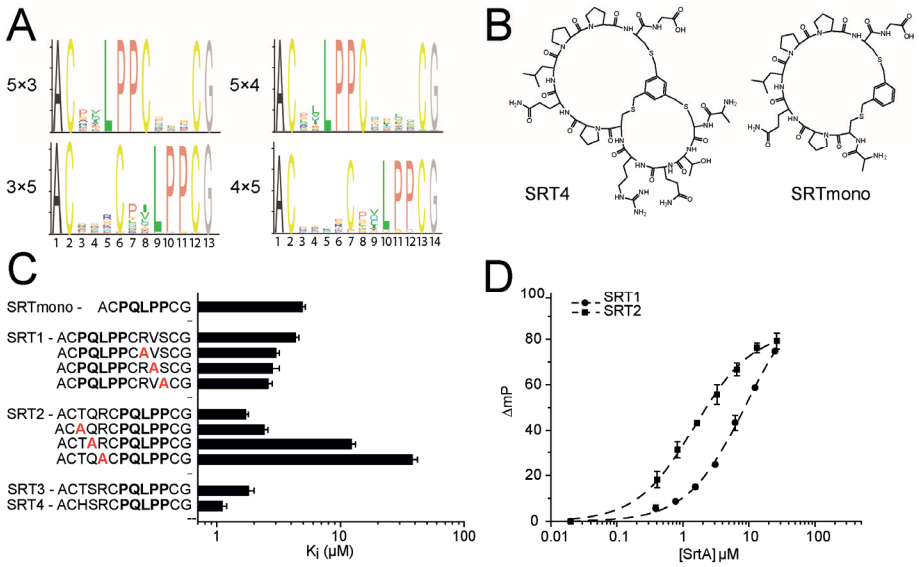


Figure 24. (A) Sequence logos of peptides containing the CxxLPPC motif in the first or in the second loop of bicyclic peptides. (B) Bicyclic peptide SRT4 (ACHSRCPQLPPCG, modified with TBMB) and monocyclic peptide SRTmono (ACPQLPPCG, modified with BBMB). (C) Alanine scanning of the non-conserved ring in the bicyclic peptides SRT1 and SRT2. The non-conserved C-terminal loop of SRT1 was not able to establish interactions with the target, whereas the non-conserved N-terminal loop in SRT2 did. Searching the HTS data for more peptides with the conserved $P^{Q}/_{L}$LPP sequence in the second ring led to the identification of SRT3 and SRT4 with around 2-fold better activities. The conserved ring alone (SRTmono) had a K_i of 4.9 μM. (D) Binding affinity of SRT1 and SRT2 measured by fluorescence polarization. Affinities correlated well with inhibitory activities (SRT1 K_d = 1.5 μM, SRT2 K_d = 8.8 μM).

To test more peptides containing the "$P^{Q}/_{L}$LPP" motif in the second ring, we searched the vast sequence data from the HTS for such peptides. The next most abundant peptides we

found were SRT3 and SRT4 and inhibited SrtA with K_i values of 1.8 and 1.1 μM. A peptide with a single ring ("SRTmono") was prepared by cyclizing ACPQLPPCG with 1,3-bis(bromomethyl)benzene (BBMB, Figure 24B). The peptide inhibited SrtA with a K_i of 4.9 μM. The best bicyclic peptide was thus around 4-fold better than the monocyclic peptide.

4.2.3 Inhibition of cell adhesion by bicyclic peptides

We evaluated the capacity of the bicyclic peptides to block adhesion of *S. aureus* strain Newman to fibrinogen. In *S. aureus*, fibrinogen adhesion is mainly driven by ClfA, a SrtA-anchored surface protein. The cells were grown until mid-exponential phase (OD_{600} = 0.5) with or without the bicyclic peptide SRT4 at concentrations up to 500 μM. Cells were added to wells of 96-well microtitre plates coated with fibrinogen (5×10^7 cells per well), and adhered cells were detected with crystal violet[173]. Unfortunately, we observed no decrease in the binding capacity of the bacteria grown in the presence of inhibitor (Figure 25A). The positive control applied in this experiment, a SrtA knockout of the *S. aureus* Newman strain (SKM12) showed reduced binding of 30%. Adhesion tests were repeated using different bicyclic peptides and a different strain and ligand protein (Cowan and fibronectin), further confirming the absence of effect (APPENDIX III-Figure S9).

We speculated that small reductions of adhesion proteins on the surface of *S. aureus* may not be detected in the cell binding assay due to multivalent binding. In order to detect small changes, we measured surface levels of protein A by ELISA. Protein A is also anchored to the peptidoglycan wall by SrtA. 2×10^7 cells of the Newman strain were coated by adhesion to wells of microtiter plates and protein A detected with an antibody. SRT4-treated cultures showed similar levels of protein A on the surface (Figure 25B).

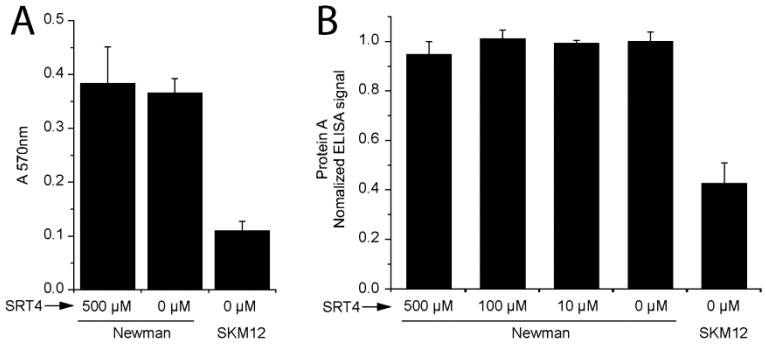


Figure 25. (A) Adhesion of *S. aureus* strains Newman to fibrinogen. The isogenic SrtA knockout strain SKM12 was used as a control. Cells were grown until OD_{600} = 0.5 in the presence or absence of 500 μM SRT4. (B) Protein A levels on the cell surface of *S. aureus* Newman (OD_{600} = 0.5) grown in the presence or absence of SRT4. SKM12 was used as a control.

To evaluate whether the lack of effect was due to degradation of the bicyclic peptides, we performed activity assays with supernatants from *S. aureus* Newman cultures after overnight incubation with bicyclic peptides. They inhibited cleavage of the fluorogenic substrate by SrtA with the expected activity. Additionally, the inhibitory activity of the bicyclic peptides was not impaired by any component in the bacterial supernatant, as the presence of 40% supernatant in the activity assay did not affect the obtained IC_{50}s (APPENDIX III-Figure S10).

4.2.4 Evaluation of reported SrtA inhibitors

It was surprising for us to find that the bicyclic peptides did not inhibit *S. aureus* adhesion while several of the small molecule inhibitors are reported inhibitors of cell adhesion. We chose to compare directly the activity of some commercially available inhibitors with the bicyclic peptides in our experimental setup.

The natural products morin and curcumin were reported to inhibit cleavage of a fluorogenic peptide substrate by SrtA both with an IC_{50} of 37.5 μM[114,121]. The same molecules were also reported to inhibit SrtB from *S. aureus* and SrtA from *S. mutans* with IC_{50}s in the same range[121,174], and to reduce biofilm formation in *S. mutans*[175,176]. We found that fluorescence of morin interfered with the fluorogenic substrate used to measure SrtA activity (Dabcyl-LPETG-edans). The cleaved SrtA substrate has fluorescence excitation and emission maxima of 335 and 495 nm (corresponding to the edans fluorophore). Morin has fluorescence excitation and emission maxima of 390 and 510 nm. Therefore in the measurements, the fluorescence was the result of the cleaved substrate and the inhibitor added to the mix. However, the fluorescence of morin decreased over time, causing the apparent slope to decrease. It is not clear how the authors accounted for this effect, which might have previously been interpreted as inhibition activity. After correction for this interference effect, no inhibition of SrtA was observed at the expected concentrations. In the case of curcumin, no interference was observed, but the inhibition activity was lower than reported (APPENDIX III-Figure S11).

We next considered the SrtA inhibitor DMMA (IC_{50} = 9.1 μM) and phenyl-vinyl sulfone (IC_{50} = 736 μM) as positive controls (Figure 11 in pg. 21). The activity of these compounds on cell adhesion had been tested with the Newman strain adhering to fibronectin-coated surfaces[124,125]. Adhesion of this strain to fibronectin is limited since it carries a truncated version of the two fibronectin-binding proteins (FnBPs) that is lacking the SrtA-recognition motif[177]. Inhibition of SrtA would therefore not have an effect on the binding of this strain to fibronectin. The observed reduced cell adhesion with the small molecules was thus most likely resulting from unspecific effects on other targets or on growth.

We anyway tested one of the compounds that was commercially available, phenyl-vinyl sulfone (PVS). PVS inhibited SrtA with a similar IC_{50} in our assay (900 μM, APPENDIX III-Figure S12). However, the inhibitor did not inhibit the binding capabilities of *S. aureus*

Newman to fibrinogen (Figure 26B). We found that PVS greatly slowed growth of the Newman strain at values close to the IC_{50} (Figure 26A). In the study that reported inhibition of cell adhesion for PVS, cell cultures with and without the inhibitor were grown in parallel and the binding of cells from these cultures was directly assessed without normalizing the cell number. It is likely that the PVS treatment led to a smaller number of cells and that this was the reason that fewer cells bound to the fibronectin coated wells. We concluded that the observed effect in decreased adherence to fibronectin by PVS was not SrtA-mediated but due to the reduced number of cells compared to the control upon addition of PVS.

The study with the reported SrtA inhibitors suggested that inhibition of *S. aureus* cell adhesion via blocking SrtA activity by small molecules is not yet established. While the observations were disillusioning, they enforced our efforts to further test if SrtA on *S. aureus* cells is inhibited by the bicyclic peptides.

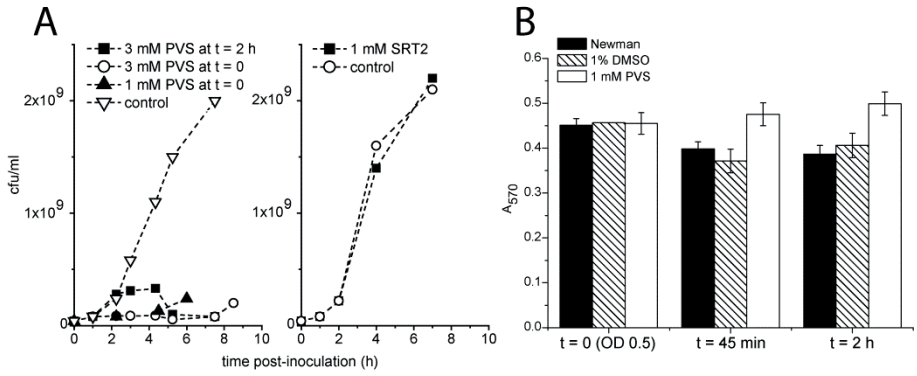


Figure 26. (A) Effect of SrtA inhibitors on growth. *S. aureus* Newman was grown in the presence or absence of PVS (left) or SRT2 (right). Although PVS has a MIC = 6 mM, it drastically slows growth at lower values. In comparison, bicyclic peptide inhibitors did not affect growth at the highest concentration tested (1 mM). (B) Adhesion of *S. aureus* Newman to fibrinogen upon treatment with PVS. 1 mM PVS (1% DMSO final concentration) was added to the culture at OD_{600} = 0.5. Aliquots were taken immediately after addition (t = 0), after 45 minutes and after 2 h.

4.2.5 Inhibition of SrtA on cells by bicyclic peptides

The activity of SrtA on *S. aureus* cells can be measured with fluorescently labeled peptides containing the LPETG sequence. Such substrates, when added to the culture, can be incorporated to the cell wall of *S. aureus* by SrtA[178,179]. We synthesized a fluorescein-labeled LPETG substrate and analyzed its incorporation to the cell wall in the presence of SRT4. SRT4 was able to prevent the incorporation of such externally added substrate, albeit with a higher IC_{50} than in the *in vitro* assays (Figure 27). This demonstrates that SRT4 was able to inhibit native SrtA on the membrane of *S. aureus*. However, it was not sufficient to prevent anchoring of the natural substrates of the enzyme. Much higher po-

tencies seem to be needed to efficiently block SrtA on the surface of *S. aureus*. A limiting factor to efficiently prevent SrtA-mediated anchoring of surface proteins could be the co-localization of native substrates and SrtA on the bacterial membrane and at specific locations (e.g. at the septum cross wall[180]), which implies that native substrates may have higher effective concentrations. A second limiting factor could be the diffusion rate of SrtA inhibitors through the cell wall, and in particular through the septum cross wall, to efficiently reach such locations.

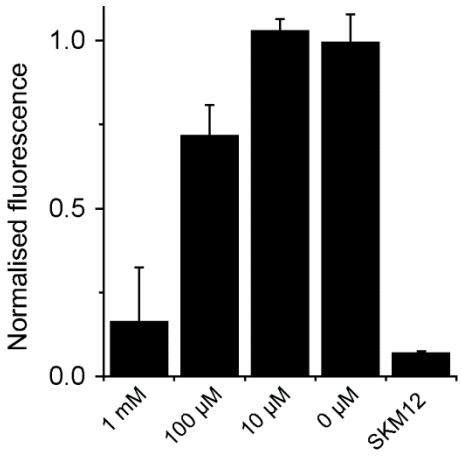


Figure 27. Inhibition of SrtA-mediated incorporation of external fluorescent substrates. *S. aureus* Newman was grown in the presence of 0.3 mM Fluo-GSLPETGGS and different concentrations of SRT4 for 24 hours. The SrtA knockout strain SKM12 was used as a control. Averages and standard deviations of at least three independent experiments are shown.

4.3 Conclusion

Since its initial discovery, SrtA has been proposed as a potential antivirulence target and while many sortase inhibitors have been described, no antisortase drugs have reached the clinic yet. Previous studies with SrtA inhibitors suggested that effective blocking of SrtA on the cells could be achieved with potencies in the micromolar range. However, certain studies have been performed using unsuitable enzymatic and adhesion assays. Moreover, the specificity of SrtA inhibitors was generally assessed by MIC values, which do not guarantee the absence of effect on growth (as in the case of PVS). In our study, we have developed selective bicyclic peptide inhibitors of SrtA with potent *in vitro* activities. They were also able to inhibit SrtA-mediated anchoring of external synthetic substrates to the cell wall, although with much lower potency, and they could not compete with the natural substrates of SrtA. More potent inhibitors are needed to effectively reach and block SrtA on *S. aureus* cells than previously thought. Bicyclic peptide inhibitors can reach nanomolar activities towards their targets, and affinity maturation of the

peptides reported in this work could lead to promising candidates for the development of antisortase therapeutics.

4.4 Experimental procedures

4.4.1 Phage selection

Libraries A, B and 6×6 were previously described[43,154]. In these libraries, peptides are displayed on five copies of the phage coat protein pIII. Libraries A and B contain peptides of the format ACX_mCX_nCG (C = cysteine, X = any amino acid). In library A, the combinations of 'm' and 'n' are 3×4, 4×3, 4×4, 3×5 and 5×3; in library B they are 3×6, 6×3, 4×5 and 5×4. Library 6×6 contains peptides of the format ACX_6CX_6CG. Phage production, reaction of cysteines with chemical linker to generate bicyclic peptides on phage, and phage panning against the different targets were performed as described before[58,62,154]. SrtA was produced as previously described[120] (APPENDIX III). SrtA was biotinylated and immobilized on streptavidin magnetic beads (Dynabeads M-280, Life Technologies, Carlsbad, CA, USA) for the first round, and on neutravidin magnetic beads for the second round. High throughput sequencing was performed using Ion Torrent PGM™, and is described in APPENDIX III.

4.4.2 Peptide synthesis

Peptides were synthesized in house by standard solid-phase peptide synthesis using Fmoc-protected amino acids (scale 0.03 mmol). As solid support, Rink amide AM resin was used to obtain peptides with a free N-terminus and an amidated C-terminus. Peptides were cleaved from the resin under reducing conditions (90% TFA, 2.5% H_2O, 2,5% thioanisol, 2,5% phenol, 2.5% 1.2-ethanedithiol) and partially purified by precipitation. In the case of bicyclic peptides, crude peptide at 0.5 mM was reacted with 1 mM TBMB in 80% aqueous buffer (20 mM NH_4HCO_3, 5 mM EDTA, pH 8.0) and 20% acetonitrile for 1 h at 30 °C. The product was purified by RP-HPLC on a C18 column, and H_2O/0.1% TFA and 95% ACN/5% H_2O/0.1% TFA were used as solvents. Pure bicyclic peptides were lyophilized and dissolved in water. The purity was assessed by RP-HPLC and was >95% for all peptides. The identity was confirmed by ESI or MALDI-TOF spectrometry.

Fluorescein-labelled peptides SRT1 and SRT2 were labeled by incubating 1 mM peptide with 3 mM 5(6)-carboxyfluorescein N-hydroxysuccinimide ester (Sigma-Aldrich, St. Louis, USA) in 40 µL PBS for 3 hrs at RT. 0.96 mL H_2O containing 0.1% TFA (v/v) was added to the reaction mixture and the peptide purified by HPLC on an analytical C18 column (Vydac C18, 218TP column, 4.6×250 mm) using a solvent system of 99.9% H_2O/0.1% TFA and 99.9% ACN/0.1% TFA. The fluorescein-modified peptide was lyophilized and the mass confirmed by ESI-MS.

For the substrate Fluo-GSLPETGGS, 5(6)-carboxyfluorescein (Sigma–Aldrich, St. Louis, USA) was coupled to the N-terminus of the peptide GSLPETGGS during solid-phase peptide synthesis. 2 equiv (0.06 mmol) of 5(6)-carboxyfluorescein, HOBt, and DCC, each in 0.25 ml DMF were added to the resin and incubated for 30 min at 400 rpm. The resin was washed four times with DMF and the fluorescence-labeled peptides cleaved as described above.

4.4.3 *In vitro* SrtA activity assays

Inhibitory activity of bicyclic peptides was determined by incubation of 2.5 μM SrtA with various peptide concentrations and quantification of the residual activity with 20 μM of fluorogenic substrate Dabcyl-LPETG-Edans (Anaspec, Fremont, USA) and 200 μM of triglycine (Sigma-Aldrich, St. Louis, USA). Residual enzymatic activities were measured in reaction buffer (50 mM Tris-Cl, pH 7.4, 150 mM NaCl, 5 mM $CaCl_2$) containing 0.1% w/v BSA in a volume of 75 μL. Fluorescence intensity was measured with a multiwell plate reader (Infinite® 200 PRO, TECAN, Maennedorf, Switzerland) (excitation at 350 nm, emission at 480 nm). The reactions were performed at 37 °C. The inhibitory constant K_i was calculated using equations described in the APPENDIX III, and was not influenced by the presence of triglycine (data not shown). Average and standard deviations of at least three measurements are reported. Phenyl vinyl sulfone, curcumin and morin hydrate (Sigma-Aldrich, St. Louis, USA) were used as received without further purification. Activity assays with these inhibitors were performed as described above including 1.5% DMSO in the reaction buffer (which did not affect the activity of the enzyme). In the case of curcumin, reaction buffer without BSA was used.

4.4.4 Fluorescence polarization

SrtA was serially diluted in reaction buffer. Fluorescein-labeled peptide at a concentration of 200 nM was also prepared in reaction buffer. 30 μL of each SrtA and fluorescein-peptide solutions were transferred into a well of a black 96-well half area microplate (Greiner Bio-One international AG, Monroe, USA) and incubated at room temperature for at least 15 min. The fluorescence polarization of each solution was measured in a multiwell plate reader (Infinite® 200 PRO, TECAN, Maennedorf, Switzerland) using a 485 nm excitation filter and a 535 nm emission filter. The dissociation constants (K_d) were determined by non-linear regression analyses of fluorescence polarization (Fp) versus total concentration of SrtA using the equation described in APPENDIX III.

4.4.5 *S. aureus* growth

S. aureus SrtA knockout strain SKM12 was kindly provided by Prof. Olaf Schneewind (University of Chicago, IL, USA). For all assays, *S. aureus* was grown in TSB medium at 37 °C, SKM12 was grown in TSB medium supplemented with 50 μg/mL erythromycin.

For adhesion and protein A ELISA tests, cultures were grown until mid-exponential phase (OD_{600} = 0.5) in the presence or absence of inhibitor, cells were pelleted and washed 3 times with cold PBS, and an aliquot was taken for cell count determination. Washed pellets were stored at -20 °C until use.

4.4.6 *S. aureus* adherence

Adhesion tests were performed as described previously[173]. 96-well plates (Nunc-Immuno MaxiSorp, Sigma-Aldrich) were coated with 1 µg/well fibrinogen or 1 µg/well fibronectin in PBS overnight at 4 °C. They were washed 3 times with PBS and subsequently blocked with PBS + 2% BSA for one hour at RT. *S. aureus* pellets from cultures in the presence or absence of the inhibitor at OD 0.5 were re-suspended in PBS to a final concentration of 6.25×10^8 cells/mL, 80 µL were added to each well (5×10^7 cells/well), and incubated at 37 °C without shaking for 90 minutes. After washing with PBS, the plate was dried for 45 minutes at 50 °C. 80 µL crystal violet was added to each well and incubated for 45 minutes at RT. After incubation with crystal violet, the plate was extensively washed with PBS and 80 µL citrate buffer (20 mM, pH 4.3) was added to each well. Absorbance at 570 nm was measured after 45 minutes.

4.4.7 *S. aureus* protein A ELISA assays

S. aureus pellets from cultures in the presence or absence of the inhibitor at OD_{600} = 0.5 were re-suspended in PBS to a final concentration of 2.5×10^8 cells/mL. 96-well plates (Nunc-Immuno MaxiSorp, Sigma-Aldrich) were coated with 2×10^7 cells/well for 90 minutes at RT. They were washed 3 times with PBS and subsequently blocked with PBS + 1% BSA for one hour at RT. After washing with PBS, anti-protein A HRP-conjugated antibody (ab7245, Abcam, Cambridge, UK) was added (1:30,000 in PBS-BSA) and the plate was incubated at RT for 30 minutes. The plate was extensively washed with PBS and 80 µL TMB substrate (Ultra TMB-ELISA, Thermo Scientific, Waltham, USA) was added to each well. After 10 minutes the reaction was quenched with 2 M H_2SO_4 and absorbance at 470 nm was measured. Data from different plates were normalized to the signal of the Newman control (untreated).

4.4.8 Incorporation of synthetic substrates on *S. aureus*

S. aureus Newman or SKM12 were grown in the presence of 0.3 mM Fluo-GSLPETGGS and different concentrations of SRT4. After 24 h, cells were pelleted and washed with cold PBS. Non-covalently bound molecules to the cell wall were removed by treatment with 5% SDS at 60 °C for 5 minutes. Cells were pelleted and washed again twice with cold PBS. Fluorescence of the cells was measured in a multiwell plate reader (Infinite® 200 PRO, TECAN, Maennedorf, Switzerland) using a 485 nm excitation filter and a 535 nm emission filter.

Chapter 5

Conclusion and outlook

The strategy for the generation and screening of libraries of bicyclic peptides by phage display was recently developed and is well established in our laboratory. The goals of my PhD were three-fold. Firstly, I wanted to explore ring size diversity in bicyclic peptide ligands in order to find potent inhibitors to a wider range of targets. Secondly, I aimed to apply high throughput sequencing to the outcome of the selections, in order to gain a deeper insight into sequence diversity. Thirdly, I tried to combine these improvements to identify new antibacterial molecules. In this last chapter, I summarize the findings and discuss the outlook of this work.

5.1 Ring size diversity in bicyclic peptides

Selections with an initial proof-of-concept library, consisting of bicyclic peptides having a fixed ring size of six amino acids (format 6×6), had yielded potent and selective binders to different targets. However, this ring size combination might not be optimal to identify the most potent binders to all targets. To systematically explore different ring size combinations, I generated libraries of bicyclic peptides containing all possible combinations of loops of 3, 4, 5, or 6 amino acids. The resulting structures are smaller than the 6×6 peptides and therefore more constrained molecules.

As a first trial, I performed selections against the protease uPA, for which the nanomolar inhibitor UK18 had been previously identified from the 6×6 library used initially in our laboratory. I hoped to find a more suitable ring size combination that would lead to the discovery of even more potent inhibitors. Interestingly, micromolar inhibitors from all combinations of loop lengths could be isolated, suggesting that a variety of peptide formats can be accommodated in the active site of uPA. Each ring size combination had a preference for one or several different amino acid motifs, and certain motifs were exclusively found in peptides with a defined ring size combination. Among the inhibitors isolated, the most potent ones corresponded to the 6×6 format and contained a RGR motif also present in the previously found UK18. In the case of uPA, more potent binders could not be isolated. However, the identification of diverse leads containing different motifs represents a great asset. Of these, certain motifs can inherently present disadvantages

(e.g. if they are cleaved by proteases in plasma). Moreover, some binding orientations may allow affinity maturation of non-conserved regions, helping to achieve higher potencies.

In contrast, when panning these variable ring size libraries against SrtA, there was a clear bias towards specific loop lengths, namely loops of 5 amino acids, and only one binding motif could be identified. It is noteworthy that no other consensus motif could be identified in spite of deep sequencing analysis of peptides selected after one or two rounds of phage selection. This highlights that the importance of ring size diversity varies from target to target, as some might have pockets able to accommodate peptides of many different formats, such as uPA, and some might only allow peptides of a certain format, such as SrtA.

These bicyclic peptide libraries with different ring sizes can be used for selections against any future target, and are currently being applied in our laboratory and others in different research projects. When combined with the newly available cyclization reagents developed in our laboratory, a huge repertoire of peptide macrocycles can be obtained, increasing the chances to find potent peptide leads for drug development.

5.2 HTS of bicyclic peptide selections

In a second part of my work, I applied high throughput sequencing (HTS) for the analysis of phage-selected peptides. In our laboratory, typically one or several 96-well plates of selected clones were sequenced per selection experiment, a costly and labor-intensive procedure. Sequencing the outputs by HTS would allow the coverage of all the diversity of the selected peptides. This can help to identify specific target-binding motifs and provides useful information about conserved residues for affinity maturation or for rational peptide design.

We decided to use the Ion Torrent PGM™ platform because it provides an adequate throughput and read length for our needs (5×10^6 sequences of 200 bp length on the Ion 316™ Chips used). Given the absence of broadly applicable software, I needed to develop tailor-made bioinformatic tools for data analysis and interpretation. These tools provide a complete processing pipeline for HTS data, from the initial sequencing file to the identification of conserved motifs in the dataset. They include: (i) a quality filter to remove non-reliable reads, (ii) a step to correct homopolymer sequencing errors, a relatively frequent (around 1%) error in Ion Torrent platforms, (iii) classification of the peptides according to the number of cysteine residues and ring sizes, and (iv) clustering of the sequences into similarity groups, allowing the identification of conserved motifs. These tools can be applied to peptide selections derived from different libraries and sequenced on different platforms. One limitation is the number of different peptide sequences that can be analyzed for similarity. At present, 200 sequences can be compared within minutes on a standard laptop computer, but the processing time increases

quadratically, putting an upper limit to the number of different sequences that can be compared.

In spite of these limitations, the developed procedures for Ion Torrent sequencing and data analysis represent broadly applicable tools, which are currently used by our group and others. In addition to the gain in time, it provides much more information than standard clone-picking and Sanger sequencing. The developed analysis software is also well documented and can be used on standard computers.

5.3 Bicyclic peptide inhibitors of sortase A

Concerning the development of bicyclic peptide inhibitors of the antivirulence target sortase A, many questions remain open. I developed bicyclic peptide inhibitors of sortase A from *Staphylococcus aureus* with potent *in vitro* activities. However, bicyclic peptide inhibitors against SrtA were around 100-fold less potent against the natively expressed SrtA on cell surface than against the recombinantly expressed SrtA in enzymatic *in vitro* assays. It would be interesting to evaluate whether this is also the case for other reported small molecule inhibitors. This highlights the importance of having representative *in vitro* assays of physiological scenarios for the development of successful drug leads. Limited diffusion through the cell wall, restricted accessibility to the active site and/or co-localization of the target with their natural substrates might be responsible for this effect, although further studies are needed to elucidate mechanistic details. This will provide essential information for the future design of new antisortase drugs.

Additionally, this work underlines the need of specific ligands. Previously reported small molecule inhibitors of SrtA were reported to have activities in the micromolar range and seemed to be sufficient to effectively block SrtA-mediated anchoring of virulence factors on the surface of the microorganism. However, the effects of some of them were most likely due to off-target effects. Potent and specific inhibitors are not only useful for the development of therapeutics, but also to gain a better understanding about biological processes and how to target them. In this sense, bicyclic peptide inhibitors might constitute valuable research tools. In order to efficiently block SrtA on *S. aureus* cells, more potent inhibitors are needed than previously thought. Bicyclic peptides with nanomolar or even picomolar activities have been developed in our laboratory against certain targets. It is likely that the inhibitors developed in this work can be improved by affinity maturation and rational design, and may constitute promising leads for the future development of antisortase therapeutics.

APPENDIX I. Supplementary Information for Chapter 2

Supplementary experimental procedures

Vector 21tet(5) cloning

The vector 21tet(5) is based on the phage vector fdg3p0ss21 and contains a 2.5 kb stuffer fragment instead of the gene region coding for D1 and D2 domains of phage p3. The vector was obtained by ligating two *Sfi*I-digested PCR products amplified with the primer pairs g3pNba/pelbsfifo and tetsfiba/tetsfifo from the vectors fdg3p0ss21 (around 7 kb) and fd-tet-DOG1[181](around 2.5 kb). The ligated DNA was transformed into TG1 cells.

Production and biotinylation of human uPA

The catalytic domain of human uPA (a N145Q mutant deficient in the glycosylation site) was expressed as pro-enzyme (pro-uPA) in mammalian cells and purified as previously described[44]. After the purification, pro-uPA showed an apparent molecular mass of about 32 kDa in SDS-PAGE. The protein was subsequently activated by plasmin cleavage: 12.3 mg pro-uPA were incubated with 49.4 µg plasmin (HPLM, from human plasma, 85 kDa, Molecular Innovations, Novi, MI, USA) in 50 mM HEPES, 150 mM NaCl, pH 8 for 6 hours at room-temperature. The activated protein (uPA) was purified by size exclusion chromatography using a HiLoad 16/60 Superdex 75 prep-grade column (GE Healthcare, Glattbrugg, Switzerland). The protein was eluted as a monomer giving a single band in SDS-PAGE, confirming the complete cleavage, with a molecular mass of about 28 kDa under reducing conditions (see Figure S1).

For biotinylation, uPA (10 µM) was incubated with EZ-link Sulfo-NHS-LC-biotin (200 µM; Pierce) in 50 µl PBS (pH 8) for 1 h at 25 °C. Excess of biotinylation reagent was removed by gel filtration with a Sephacryl S100 column (GE Healthcare) using 50 mM NaAc buffer, 200 mM NaCl, pH 5.5. The ability of the biotinylated uPA to bind to either streptavidin or neutravidin was verified by incubating the protein with magnetic streptavidin and neutravidin beads respectively and analyzing the bound and unbound protein fraction by SDS-PAGE.

Chemical synthesis of peptides

Fmoc-protected amino acids and Fmoc-rink amide AM resin (0.26 mmol/g resin) were purchased from Iris Biotech GmbH (Marktredwitz, Germany). O-Benzotriazole-N,N,N',N'-tetramethyl-uronium-hexafluoro-phosphate (HBTU, ChemPep, Wellington, FL, USA), N,N-Diisopropylethylamine (DIPEA, Merck Schuchardt OHG, Hohenbrunn, Germany), trifluoroacetic acid (TFA, Sigma-Aldrich Chemie GmbH, Steinheim, Germany), 1,2-

ethanedithiol (EDT, Fluka Chemie GmbH, Buchs, Switzerland), thioanisole (Fluka), piperidine (Fluka), phenol (Acros Organics, Geel, Belgium) and 1,3,5-tris(bromomethyl)benzene (TBMB) were used as received without further purification. Peptides were synthesized on an Advanced ChemTech 348Ω peptide synthesizer (Aapptec, Louisville, USA) by standard Fmoc (9-fluorenylmethyloxycarbonyl) solid-phase chemistry on a Rink Amide AM resin (0.03 mmol scale). Coupling steps were carried out twice, for each time amino acid (4 eq., 0.2 M solution in DMF), HBTU (4 eq, 0.45 M solution in DMF), OxymaPure (4 eq, 0.45 M solution in DMF) and DIPEA (6 eq, 0.5 M solution in DMF) were used. Fmoc groups were removed using a 20% (v/v) solution of piperidine in DMF (2.5 ml × 2). The final peptides were deprotected (side-chain protected groups) and cleaved from the resin using a TFA/thioanisole/H_2O/phenol/EDT mixture (90/2.5/2.5/2.5/2.5 v/v, 4 ml) for 3 hours at room temperature. The resin was removed by filtration under vacuum and the peptides were precipitated with cold diethyl ether (40 ml). The precipitated peptides were resuspended and washed twice with diethyl ether (20 ml each time). Finally, the peptides were dissolved in H_2O: CH_3CN (1:1) and lyophilized.

Mass spectrometric analysis of synthetic peptides

The molecular masses of synthetic peptides before and after chemical modification were determined with an Axima-CFR plus MALDI-TOF mass spectrometer (Axima-CFR plus, Kratos Shimadzu Biotech, Manchester, UK). HPLC-purified peptides (0.1-10 μM in 0.1% v/v TFA/10-30% v/v CH_3CN in water) were mixed 1:1 with a saturated solution of matrixα-cyano-4-hydroxycinnamic acid (α-CHCA) in 50% v/v CH_3CN, 49.9% v/v H_2O, 0.1% v/v TFA and loaded onto a MALDI carrier plate for mass determination.

Supplementary table

Primer name	DNA sequence
g3pNba	5'-CAGTCAGGCCTCGGGGGCCATGGCTTCTGGTACCCCGGTTAAC-3'
Pelbsfifo	5'-GACTGAGGCCGGCTGGGCCGCATAGAAAGGAACAACTAAAGGAAT-3'
Tetsfiba	5'-CAGTCAGGCCCAGCCGGCCGATCTCGGGAAAAGCGTTGGTCAC-3'
Tetsfifo	5'-GACTGAGGCCCCCGAGGCCTTCCCTTTGTCAACAGCAATGG-3'
Prepcr	5'-GGCGGTTCTGGCGCTGAAACTGTTGAAAGTAG-3'
sfi2notfo	5'-CCATGGCCCCCGAGGCCGCGGCCGCATTGACAGG-3'
SfiIcx3cx4cba	5'-TATGCGGCCCAGCCGGCCATGGCAGCATGCNNKNNKNNKTGCNNKNNKNNKNNKTGTGGCGGTTCTGGCGCTG-3'
SfiIcx4cx3cba	5'-TATGCGGCCCAGCCGGCCATGGCAGCATGCNNKNNKNNKNNKTGCNNKNNKNNKTGTGGCGGTTCTGGCGCTG-3'
SfiIcx3cx5cba	5'-TATGCGGCCCAGCCGGCCATGGCAGCATGCNNKNNKNNKTGCNNKNNKNNKNNKNNKTGTGGCGGTTCTGGCGCTG-3'
SfiIcx4cx4cba	5'-TATGCGGCCCAGCCGGCCATGGCAGCATGCNNKNNKNNKNNKTGCNNKNNKNNKNNKTGTGGCGGTTCTGGCGCTG-3'
SfiIcx5cx3cba	5'-TATGCGGCCCAGCCGGCCATGGCAGCATGCNNKNNKNNKNNKNNKTGCNNKNNKNNKTGTGGCGGTTCTGGCGCTG-3'
SfiIcx3cx6cba	5'-TATGCGGCCCAGCCGGCCATGGCAGCATGCNNKNNKNNKTGCNNKNNKNNKNNKNNKNNKTGTGGCGGTTCTGGCGCTG-3'
SfiIcx4cx5cba	5'-TATGCGGCCCAGCCGGCCATGGCAGCATGCNNKNNKNNKNNKTGCNNKNNKNNKNNKNNKTGTGGCGGTTCTGGCGCTG-3'
SfiIcx5cx4cba	5'-TATGCGGCCCAGCCGGCCATGGCAGCATGCNNKNNKNNKNNKNNKTGCNNKNNKNNKNNKTGTGGCGGTTCTGGCGCTG-3'
SfiIcx6cx3cba	5'-TATGCGGCCCAGCCGGCCATGGCAGCATGCNNKNNKNNKNNKNNKNNKTGCNNKNNKNNKTGTGGCGGTTCTGGCGCTG-3'
SfiIcx4cx6cba	5'-TATGCGGCCCAGCCGGCCATGGCAGCATGCNNKNNKNNKNNKTGCNNKNNKNNKNNKNNKNNKTGTGGCGGTTCTGGCGCTG-3'
SfiIcx5cx5cba	5'-TATGCGGCCCAGCCGGCCATGGCAGCATGCNNKNNKNNKNNKNNKTGCNNKNNKNNKNNKNNKTGTGGCGGTTCTGGCGCTG-3'
SfiIcx6cx4cba	5'-TATGCGGCCCAGCCGGCCATGGCAGCATGCNNKNNKNNKNNKNNKNNKTGCNNKNNKNNKNNKTGTGGCGGTTCTGGCGCTG-3'
SfiIcx5cx6cba	5'-TATGCGGCCCAGCCGGCCATGGCAGCATGCNNKNNKNNKNNKNNKTGCNNKNNKNNKNNKNNKNNKTGTGGCGGTTCTGGCGCTG-3'
SfiIcx6cx5cba	5'-TATGCGGCCCAGCCGGCCATGGCAGCATGCNNKNNKNNKNNKNNKNNKTGCNNKNNKNNKNNKNNKTGTGGCGGTTCTGGCGCTG-3'

Table S1. DNA sequences of the primers (5' to 3') used for 21tet(5) vector generation and library cloning. *Sfi*I restriction sites are underlined. Diversity was introduced by using the degenerate codon NNK in the synthetic primers. N represents any of the 4 nucleotides and K thymidine (T) and guanosine (G).

Supplementary figures

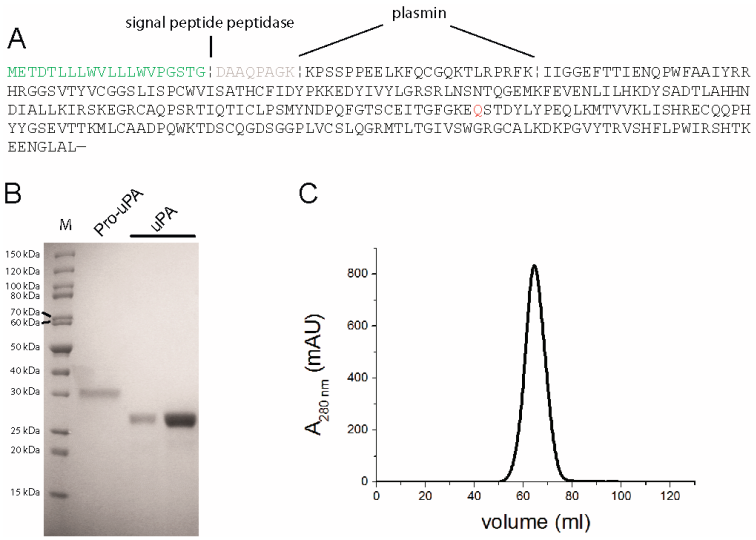


Figure S1. Expression and purification of the catalytic domain of human uPA. (A) Amino acid sequence of the protein expressed in mammalian cells. The Ig κ-chain leader sequence is highlighted in green, 8 random amino acids inserted due to the cloning strategy are shown in grey and the C-terminal fragment of chain A as well as the catalytic domain of human uPA (chain B) are shown in black wherein a mutated residue (N145Q) is highlighted in red. Signal peptide peptidase and plasmin proteolysis cleavage sites are indicated by black broken lines. (B) SDS-PAGE analysis of the protein secreted by mammalian cells before (Pro-uPA, approximately 31 kDa) and after activation with human plasmin (uPA, approximately 28 kDa) under reducing conditions. The protein was stained with Coomassie blue dye. M: molecular weight marker. (C) Active uPA-N145Q (uPA), analyzed by size exclusion chromatography, showed a high degree of purity.

Name	Peptide sequence	m/z expected (TBMB modified)	m/z found (TBMB modified)
UK327	ACTARTCPATQVLCG	1607.7	1608.0
UK339	ACNWKFSLCETQRNQCG	2100.9	2101.6
UK340	ACNSRFALCSPSSQMCG	1874.8	1875.3
UK343	ACTEFQTDCRGRSSICG	1946.8	1948.3
UK344	ACNHAATDCRGRGGPCG	1758.7	1759.0
UK346	ACKQSVCTARTLCG	1553.7	1554.5
UK348	ACKHSDCTARFPCG	1608.7	1608.9
UK368	ACRGGCKFTMCG	1346.6	1346.8
UK377	ACLQGERGCENRRPSCG	1948.9	1948.9

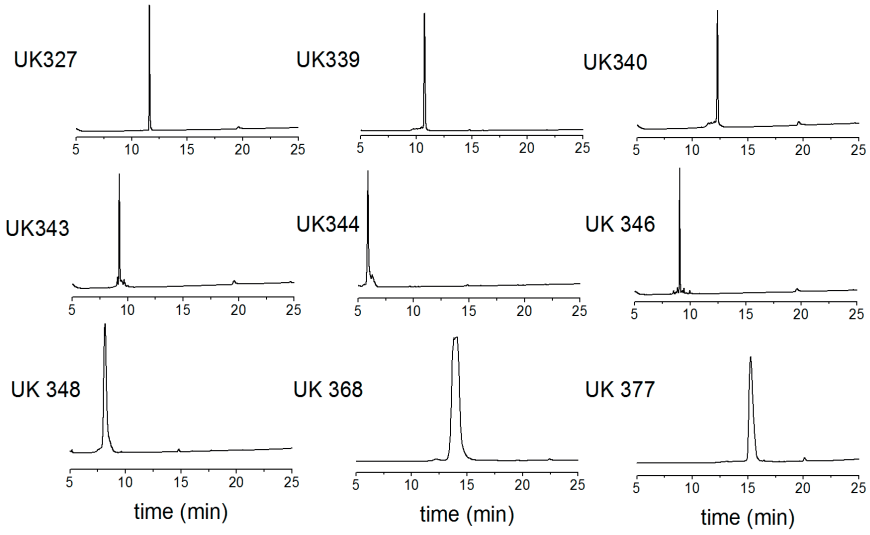


Figure S2. Sequence, mass values predicted and experimentally determined by MALDI-TOF (top) and analytic HPLC chromatograms (bottom) for selected peptides after TBMB modification.

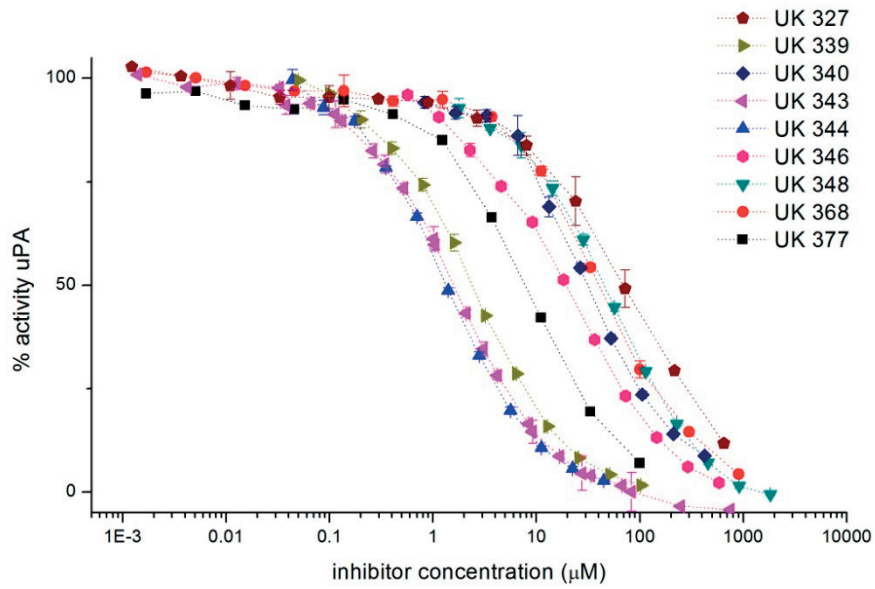


Figure S3. Inhibition of human uPA by the nine TBMB modified peptides tested. The residual activity of the protease was measured with a fluorogenic substrate and plotted against the concentration of bicyclic peptide (logarithmic scale).

APPENDIX II. Supplementary Information for Chapter 3

Supplementary results

Barcode assignment

Barcodes were designed so they could be identified even if one or two of the six bases were wrongly sequenced. We included in the MatLab script an option to allow one mismatch in the barcode (one insertion, mutation or deletion). Application of this procedure did not increase much the number of sequences that could be used because most of the rescued sequences were filtered out in the subsequent quality filter due to bad quality values in the peptide region (Figure S4). For all the analysis in this work, we therefore used only sequences in which the barcode showed a perfect match.

Validation of the mathematical model to estimate the number of different sequences

For a homogeneous population of sequences (where all clones are equally represented in the pool), the number of different sequences found (y) in function of the number of sequences sampled (x) will increase linearly at the beginning, decreasing the rate as it approaches saturation (Figure S5). This system could be approximated by equation S1

$$(a - y) = \mathrm{k}\frac{dy}{dx} \qquad \text{(S1)}$$

where a is the total number of different sequences in the pool, and the rate of finding new sequences (dy/dx) is proportional to the number of new sequences remaining in the sample, with a proportionality constant k. Solving the differential equation, the number of different sequences corresponds to equation S2:

$$y = a\left(1 - e^{-x/k}\right) \quad \text{(S2)}$$

Taylor approximation near 0, where the function behaves almost linearly (x << a), allows the determination of the initial slope of the curve as shown in equation S3:

$$y = \frac{a}{k}x, \ for\ x \ll a \qquad \text{(S3)}$$

In the case of a homogeneous dataset, the initial slope should be close to one, and therefore $a = k$. Indeed, simulation of an ideal homogeneously distributed dataset gave the expected curve and fitted parameters corresponded to the ones simulated (Figure S5).

In the case of non-homogeneous datasets, where a few sequences might represent a significant fraction of the population, we anticipated that the system would behave similarly, but with a lower initial slope. We validated this approach by fitting a series of simulated datasets representing populations with different abundance distributions (Figure S6). The parameter a/k takes values between 0 and 1, and could be used to quantify the homogeneity of the sample:

$$\text{(less homogeneous) } 0 < a/k \leq 1 \text{ (homogeneous)}$$

In the experimental datasets, at larger number of reads, the number of different sequences increased linearly and did not converge to a maximal value. The linear increase was due to sequencing errors, which were directly proportional to the number of reads. To give account for this effect, a linear component was added to equation S2:

$$y = a\left(1 - e^{-x/k}\right) + bx \qquad \text{(S4)}$$

where b is a global error rate for the population. Equation S4 was used to fit the data in this study. We additionally simulated the same sets presented in Figure S6, adding different percentages of random mutations, and using equation S4 to fit the data, obtaining a good estimation of the parameters (Figure S7).

Supplementary Tables

Primer name	Sequence
	Primers for PCR amplification of Library A and Library B
IT_Fw1	5' CCATCTCATCCCTGCGTGTCTCCGACTCAGGCATAGTTTCTATGCGGCCCAGC 3'
IT_Fw2	5' CCATCTCATCCCTGCGTGTCTCCGACTCAGCGTATCTTTCTATGCGGCCCAGC 3'
IT_Fw3	5' CCATCTCATCCCTGCGTGTCTCCGACTCAGATCGCATTTCTATGCGGCCCAGC 3'
IT_Fw4	5' CCATCTCATCCCTGCGTGTCTCCGACTCAGACGATATTTCTATGCGGCCCAGC 3'
IT_Fw5	5' CCATCTCATCCCTGCGTGTCTCCGACTCAGAGACTCTTTCTATGCGGCCCAGC 3'
IT_Fw6	5' CCATCTCATCCCTGCGTGTCTCCGACTCAGGATACATTTCTATGCGGCCCAGC 3'
IT_Fw7	5' CCATCTCATCCCTGCGTGTCTCCGACTCAGCATCTCTTTCTATGCGGCCCAGC 3'
IT_Fw8	5' CCATCTCATCCCTGCGTGTCTCCGACTCAGGTTCAGTTTCTATGCGGCCCAGC 3'
IT_Fw9	5' CCATCTCATCCCTGCGTGTCTCCGACTCAGTACCAGTTTCTATGCGGCCCAGC 3'
IT_Fw10	5' CCATCTCATCCCTGCGTGTCTCCGACTCAGATGGAGTTTCTATGCGGCCCAGC 3'
IT_Rev1	5' CCTCTCTATGGGCAGTCGGTGATGTTTCAGCGCCAGAACC 3'
	Primers for PCR amplification of Library 3×3 and Library 4×4
IT_Fw11	5' CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTTACCGCAATTCCTTTAGTTGTTC 3'
IT_Fw12	5' CCATCTCATCCCTGCGTGTCTCCGACTCAGGGTGAACGCAATTCCTTTAGTTGTTC 3'
IT_Rev2	5' CCTCTCTATGGGCAGTCGGTGATTTTCAACAGTTTCAGCGGAGTG 3'

Table S2. Primers for PCR amplification and subsequent Ion Torrent sequencing. Forward primers contain adaptor sequence, barcode (underlined) and template-specific sequence. Reverse primers contain adaptor sequence and template-specific sequence.

	before correcting sequencing errors			after correcting sequencing errors		
Dataset	a (# different sequences)	a/k (homogeneity)	b (error rate)	a (# different sequences)	a/k (homogeneity)	b (error rate)
SrtA – Library A	2817	0.743	3.3%	2814	0.712	1.8%
SrtA – Library B	1442	0.212	2.8%	1422	0.222	1.4%
uPA – Library B	3129	0.319	5.1%	2980	0.321	1.6%
FXII – 4×4	7884	0.671	5.1%	7839	0.719	1.1%
PK – 3×3, 4×4	1376	0.670	4.5%	1333	0.704	0.6%
SA – 3×3, 4×4	343	0.641	2.8%	311	0.725	0.2%

Table S3. Correction of errors in sequences of phage-selected peptides. Diversity, homogeneity and error rate were estimated before and after correcting sequencing errors for different datasets (datasets were obtained after one round of phage selection and the correction was applied to all sequences). The estimated error rate (parameter *b*) is considerably reduced, while the estimated number of different sequences and homogeneity remain almost unchanged.

Supplementary Figures

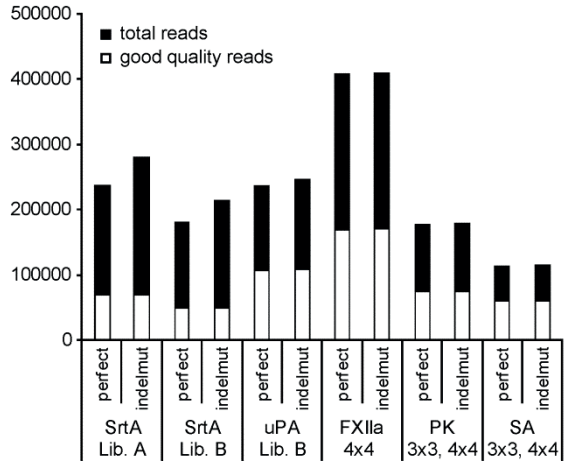


Figure S4. Number of sequences found for each barcode. Bars labeled with 'perfect' show the number of sequences having a perfect match of the barcode. Bars labeled with 'indelmut' show the sum of sequences with a perfect match and those having one insertion, deletion or mutation in the barcode. The white area within the bars shows sequences that passed the quality filter after analyzing the peptide region (quality parameters: maximum 3 bases < Q18).

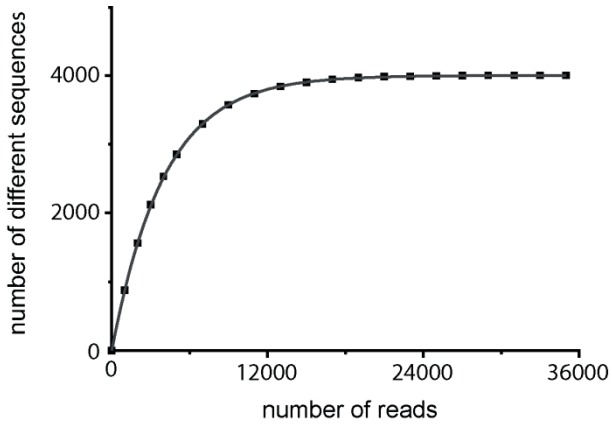
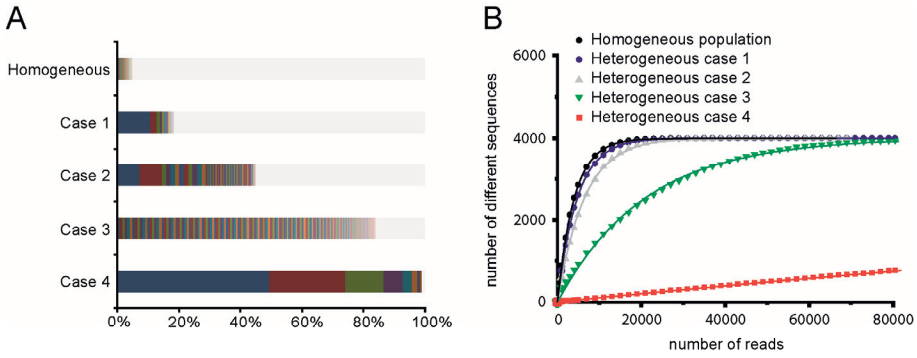


Figure S5. Simulation of an ideal dataset of homogeneously distributed sequences. The dataset was chosen to contain 4000 different sequences. The number of different sequences is indicated in dependence of the number of reads sampled (black squares). Equation 2 was used to fit the data (grey line). Calculated parameters (total number of different sequences = 4000, a/k = 0.995) corresponded with the ones simulated.

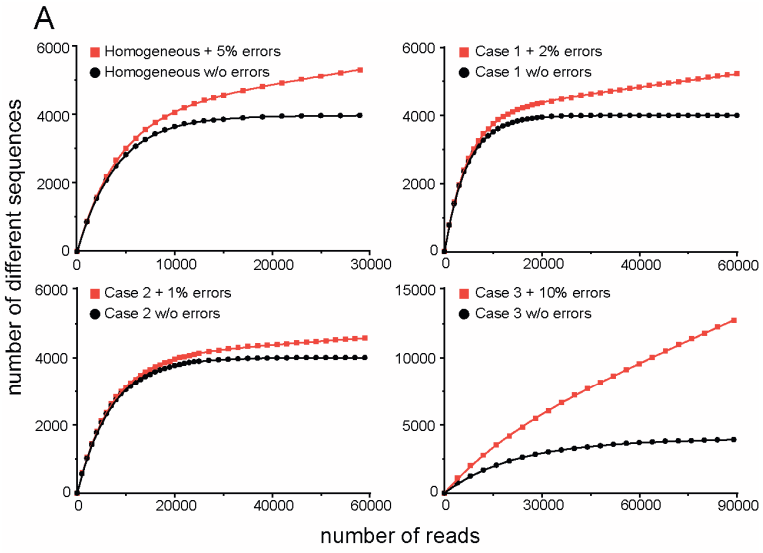


C

dataset	a (total different sequences)	a/k (homogeneity)
Homogeneous	4000	0.995
Case 1	4001	0.843
Case 2	3995	0.572
Case 3	4012	0.191
Case 4	4137	0.01

Figure S6. Simulation of datasets presenting different abundance distributions of peptide sequences. (A) Representation of the homogeneous population and different simulated heterogeneous populations (case 1 to case 4). All contained a total of 4000 different sequences but they were present in different relative abundances. Top 200 most abundant sequences are separated as blocks shown in different colors. (B) Saturation plots of these heterogeneous populations and subsequent fitting of equation S2. (C) Calculated parameters from the fitting. For case 4, 10^6 reads were sampled to reach saturation.

A



B

dataset	a (total different sequences)	a/k (homogeneity)	b (error rate)
Homogeneous + 5% errors	4005	0.961	0.0464
Case 1 + 2% errors	4052	0.854	0.0195
Case 2 + 1% errors	3995	0.572	0.0098
Case 3 + 10% errors	3411	0.165	0.106

Figure S7. Simulation of populations containing 4000 different sequences and presenting different abundance distributions and different sequencing error rates. (A) Saturation plots of simulated populations and fitting of equation S4. (B) Calculated parameters from the fitting. Good estimates for the total number of different sequences a, the error rate b and the homogeneity of the population a/k were obtained.

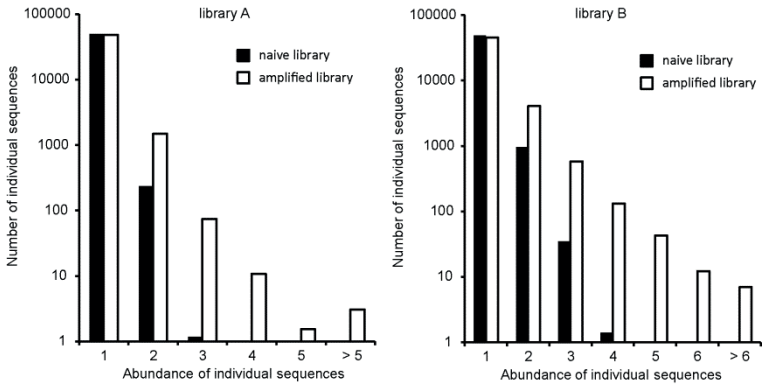


Figure S8. Libraries analyzed before (naive library) and after one round of amplification (infection of bacteria without affinity selection; amplified library). A sub-population of the library is being preferentially amplified, but the most abundant clone represents less than 0.02% of the population.

Similartiy clusters (Figure 20)

Computational comparison of peptide sequences for the identification of target-binding motifs. Sequence data was analyzed before with the MatLab scripts *Step1.m* and *Step2.m* (using standard quality parameters). Sequencing errors of the 500 most abundant sequences were corrected by script *fixingerrors.m*. Groups are shown as found by the software (stringency = 0.5 except for uPA and FXII where stringency = 0.4). The same data is shown in Figure 20 after manual editing.

SrtA – Library A

Group1		
MAACSILPPCTIQCGGSG	76	ATGGCAGCATGCTCTATTCTTCCTCCGTGCACGATTTAGTGTGGCGGTTCTGGCG
MAACSILPPCTTHCGGSG	70	ATGGCAGCATGCTCCATTTTGCCGCCTTGCACGACGCATTGTGGCGGTTCTGGCG
MAACSVLPPCSFSCGGSG	148	ATGGCAGCATGCTCTGTTCTTCCTCCCTGCTCCTTCTCGTGTGGCGGTTCTGGCG
MAACSVLPPCSVPCGGSG	57	ATGGCAGCATGCAGTGTCCTGCCTCCGTGCTCTGTCCCCTGTGGCGGTTCTGGCG
MAACSILPPCNSPCGGSG	41	ATGGCAGCATGCTCGATTCTGCCGCCGTGCAATAGTCCTTGTGGCGGTTCTGGCG
MAACGILPPCAMSCGGSG	59	ATGGCAGCATGCGGGATTCTTCCTCCGTGCGCTATGTCTTGTGGCGGTTCTGGCG
MAACSILPPCSQNCGGSG	109	ATGGCAGCATGCTCTATCCTTCCTCCTTGCTCTTAGAATTGTGGCGGTTCTGGCG
MAACALLPPCNQNCGGSG	55	ATGGCAGCATGCGCTCTGCTGCCGCCCTGCAATTAGAACTGTGGCGGTTCTGGCG
MAACTMLPPCSSNCGGSG	51	ATGGCAGCATGCACTATGTTGCCTCCTTGCTCGTCTAACTGTGGCGGTTCTGGCG
MAACPVLPPCISNCGGSG	67	ATGGCAGCATGCCCCGTTCTCCCCCCCTGCATTAGTAATTGTGGCGGTTCTGGCG
MAACTILPPCVSNCGGSG	36	ATGGCAGCATGCACCATCTTGCCGCCTTGCGTGTCCAACTGTGGCGGTTCTGGCG
MAACSILPPCFQTCGGSG	68	ATGGCAGCATGCAGCATCTTGCCCCCCTGCTTCTAGACCTGTGGCGGTTCTGGCG
MAACPSLPPCPWNCGGSG	91	ATGGCAGCATGCCCGAGTCTTCCCCCGTGCCCCTGGAATTGTGGCGGTTCTGGCG
MAACPVLPPCPINCGGSG	74	ATGGCAGCATGCCCCGTCCTGCCTCCCTGCCCTATTAACTGTGGCGGTTCTGGCG
Group2		
MAACSRSCPVLPPCGGSG	65	ATGGCAGCATGCTCTCGTTCCTGCCCTGTGCTTCCCCCCTGTGGCGGTTCTGGCG
MAACSHQCPVLPPCGGSG	35	ATGGCAGCATGCTCGCACCAGTGCCCTGTGCTCCCTCCTTGTGGCGGTTCTGGCG
MAACTNNCPLLPPCGGSG	86	ATGGCAGCATGCACTAACAATTGCCCCCTTTTGCCTCCTTGTGGCGGTTCTGGCG
MAACQTGCPILPPCGGSG	98	ATGGCAGCATGCTAGACGGGTTGCCCGATTCTGCCTCCCTGTGGCGGTTCTGGCG
MAACYTSCPVLPPCGGSG	37	ATGGCAGCATGCTACACTTCTTGCCCTGTTCTGCCTCCCTGTGGCGGTTCTGGCG
MAACGHGCPYLPPCGGSG	69	ATGGCAGCATGCGGTCACGGCTGCCCTTACCTGCCTCCTTGTGGCGGTTCTGGCG
MAACSQGCPSLPPCGGSG	43	ATGGCAGCATGCAGTTAGGGGTGCCCGTCGCTGCCTCCGTGTGGCGGTTCTGGCG
MAACVSSCPSLPPCGGSG	52	ATGGCAGCATGCGTGTCCAGTTGCCCGAGTCTTCCCCCCTGTGGCGGTTCTGGCG
MAACVSNCPYLPPCGGSG	38	ATGGCAGCATGCGTTTCTAATTGCCCCTACTTGCCGCCGTGTGGCGGTTCTGGCG
MAACISLCPQLPPCGGSG	37	ATGGCAGCATGCATTTCTCTTTGCCCTCAGCTCCCTCCTTGTGGCGGTTCTGGCG
MAACVLACPYLPPCGGSG	38	ATGGCAGCATGCGTCCTGGCCTGCCCCTATCTTCCCCCCTGTGGCGGTTCTGGCG
MAACSGLCTVLPPCGGSG	97	ATGGCAGCATGCAGTGGTTTGTGCACTGTCCTCCCTCCTTGTGGCGGTTCTGGCG
MAACRGTCPVLPPCGGSG	75	ATGGCAGCATGCAGGGGCACCTGCCCCGTTCTGCCTCCTTGTGGCGGTTCTGGCG
MAACSGYCPYLPPCGGSG	89	ATGGCAGCATGCAGCGGTTATTGCCCCTACCTCCCGCCTTGTGGCGGTTCTGGCG
MAACNGFCPSLPPCGGSG	86	ATGGCAGCATGCAATGGTTTTTGCCCGAGTCTCCCTCCTTGTGGCGGTTCTGGCG
MAACNTLCPYLPPCGGSG	146	ATGGCAGCATGCAATACTCTTTGCCCTTATCTGCCCCCTTGTGGCGGTTCTGGCG
MAACSWRCPSLPPCGGSG	194	ATGGCAGCATGCTCGTGGCGGTGCCCGTCTCTCCCTCCGTGTGGCGGTTCTGGCG
MAACSVRCDTLPPCGGSG	41	ATGGCAGCATGCAGTGTGCGGTGCGATACTCTTCCTCCCTGTGGCGGTTCTGGCG
MAACASRCHQLPPCGGSG	36	ATGGCAGCATGCGCCTCCAGGTGCCACCAGCTGCCTCCGTGTGGCGGTTCTGGCG
Group3		
MAACVLPPCFYHDCGGSG	40	ATGGCAGCATGCGTTCTCCCTCCGTGCTTTTATCATGATTGTGGCGGTTCTGGCG
MAACVLPPCFTYDCGGSG	36	ATGGCAGCATGCGTGCTCCCGCCGTGCTTTACTTATGATTGTGGCGGTTCTGGCG
MAACVLPPCSSLDCGGSG	63	ATGGCAGCATGCGTCCTGCCTCCCTGCTCCTCCTTGGACTGTGGCGGTTCTGGCG
MAACILPPCSSVDCGGSG	37	ATGGCAGCATGCATTTTGCCCCCCTGCTCGAGTGTTGATTGTGGCGGTTCTGGCG
MAACLLPPCTPLDCGGSG	89	ATGGCAGCATGCCTCCTTCCGCCCTGCACTCCCCTCGATTGTGGCGGTTCTGGCG
MAACLLPPCSYMECGGSG	37	ATGGCAGCATGCCTTCTGCCTCCTTGCAGCTATATGGAGTGTGGCGGTTCTGGCG
MAACLLPPCSYIDCGGSG	107	ATGGCAGCATGCCTCCTCCCTCCTTGCTCTTACATTGACTGTGGCGGTTCTGGCG
MAACLLPPCSYIDCGGSG	81	ATGGCAGCATGCCTTCTTCCCCCTTGCTCCTACATCGATTGTGGCGGTTCTGGCG
MAACLLPPCNFLDCGGSG	52	ATGGCAGCATGCCTCCTTCCGCCTTGCAATTTTCTGGATTGTGGCGGTTCTGGCG
MAACILPPCQFKDCGGSG	44	ATGGCAGCATGCATTCTGCCTCCCTGCTAGTTCAAGGATTGTGGCGGTTCTGGCG
MAACVLPPCTFADCGGSG	35	ATGGCAGCATGCGTTCTTCCCCCTTGCACCTTTGCGGACTGTGGCGGTTCTGGCG
MAACILPPCSYTQCGGSG	78	ATGGCAGCATGCATCTTGCCTCCCTGCTCTTACACGTAGTGTGGCGGTTCTGGCG
MAACHLPPCSLHLCGGSG	125	ATGGCAGCATGCCATCTTCCCCCTTGCTCCTTGCACTTGTGTGGCGGTTCTGGCG
MAACHLPPCDATLCGGSG	60	ATGGCAGCATGCCACCTGCCGCCTTGCGATGCTACGCTTTGTGGCGGTTCTGGCG
Group4		
MAACPTLPPCEAPCGGSG	72	ATGGCAGCATGCCCCACTCTGCCCCCTTGCGAGGCCCCGTGTGGCGGTTCTGGCG
MAACATLPPCQAPCGGSG	56	ATGGCAGCATGCGCGACTCTCCCTCCTTGCTAGGCTCCTTGTGGCGGTTCTGGCG
MAACAYLPPCEANCGGSG	56	ATGGCAGCATGCGCGTATTTGCCCCCTTGCGAGGCTAATTGTGGCGGTTCTGGCG

```
MAACLQLPPCSSPCGGSG          92      ATGGCAGCATGCTTGTAGCTTCCTCCCTGCTCGTCCCCGTGTGGCGGTTCTGGCG
MAACRQLPPCSDPCGGSG          59      ATGGCAGCATGCCGTTAGTTGCCTCCCTGCAGCGATCCCTGTGGCGGTTCTGGCG
MAACAELPPCNNPCGGSG          41      ATGGCAGCATGCGCGGAGCTGCCTCCTTGCAACAATCCTTGTGGCGGTTCTGGCG

 Group5
MAACFFPCGVLPPCGGSG          74      ATGGCAGCATGCTTCTTTCCGTGCGGTGTGCTTCCCCCTTGTGGCGGTTCTGGCG
MAACLYPCGVLPPCGGSG          50      ATGGCAGCATGCCTGTACCCGTGCGGTGTGCTGCCCCCCTGTGGCGGTTCTGGCG
MAACDFTCGILPPCGGSG          67      ATGGCAGCATGCGATTTTACTTGCGGTATTCTGCCTCCGTGTGGCGGTTCTGGCG

 Group6
MAACNIQSCLPPCGGSG          75      ATGGCAGCATGCAACATTTAGTCTTGCCTTCCTCCTTGTGGCGGTTCTGGCG
MAACTIESCLPPCGGSG          53      ATGGCAGCATGCACGATTGAGTCTTGCCTTCCTCCTTGTGGCGGTTCTGGCG
MAACSLVSCLPPCGGSG          55      ATGGCAGCATGCAGTCTCGTTAGTTGCTTGCCTCCCTGTGGCGGTTCTGGCG
MAACSIITCLPPCGGSG          45      ATGGCAGCATGCTCCATCATCACGTGCTTGCCTCCTTGTGGCGGTTCTGGCG
MAACSYELCLPPCGGSG          37      ATGGCAGCATGCAGTTATGAGTTGTGCCTCCCTCCTTGTGGCGGTTCTGGCG
MAACPFVPCLPPCGGSG          39      ATGGCAGCATGCCCTTTCGTCCCTTGCCTTCCTCCTTGTGGCGGTTCTGGCG

 Group7
MAACPILPPCHAHCGGSG          135     ATGGCAGCATGCCCCATTCTTCCCCCTTGCCATGCTCATTGTGGCGGTTCTGGCG
MAACPILPPCHIMCGGSG          38      ATGGCAGCATGCCCTATTTTGCCTCCGTGCCACATCATGTGTGGCGGTTCTGGCG
MAACPALPPCHSDCGGSG          45      ATGGCAGCATGCCCTGCGCTGCCCCCGTGCCATTCCGATTGTGGCGGTTCTGGCG
MAACPSLPPCNQHCGGSG          48      ATGGCAGCATGCCCCTCTCTCCCTCCCTGCAATTAGCATTGTGGCGGTTCTGGCG
MAACPYLPPCNGHCGGSG          43      ATGGCAGCATGCCCGTATCTGCCTCCTTGCAATGGTCATTGTGGCGGTTCTGGCG
MAACPLLPPCSLDCGGSG          317     ATGGCAGCATGCCCCCCTGCTGCCGCCTTGCAGTCTGGATTGTGGCGGTTCTGGCG
MAACPFLPPCSLKCGGSG          43      ATGGCAGCATGCCCGTTTCTTCCTCCGTGCTCTTTGAAGTGTGGCGGTTCTGGCG
MAACPLLPPCGIGCGGSG          79      ATGGCAGCATGCCCTCTGCTGCCTCCGTGCGGTATTGGCTGTGGCGGTTCTGGCG
MAACPLLPPCADDCGGSG          701     ATGGCAGCATGCCCCTTGCTCCCTCCCTGCGCTGATGATTGTGGCGGTTCTGGCG
MAACPYLPPCGTICGGSG          71      ATGGCAGCATGCCCCTATCTGCCTCCTTGCGGCACGATCTGTGGCGGTTCTGGCG
MAACPLLPPCSGICGGSG          42      ATGGCAGCATGCCCCCTCCTGCCTCCCTGCTCTGGTATCTGTGGCGGTTCTGGCG
MAACPSLPPCINSCGGSG          67      ATGGCAGCATGCCCCTCCCTGCCTCCCTGCATTAACTCGTGTGGCGGTTCTGGCG
MAACPVLPPCLASCGGSG          62      ATGGCAGCATGCCCTGTGCTCCCGCCCTGCCTGGCTTCTTGTGGCGGTTCTGGCG
MAACPRLPPCSSSCGGSG          36      ATGGCAGCATGCCCGCGTTTGCCTCCTTGCTCGAGTTCTTGTGGCGGTTCTGGCG
MAACPQLPPCRVSCGGSG          114     ATGGCAGCATGCCCTTAGCTCCCTCCCTGCCGCGTGTCTTGTGGCGGTTCTGGCG
MAACPALPPCQLSCGGSG          63      ATGGCAGCATGCCCTGCGTTGCCCCCGTGCTAGTTGTCTTGTGGCGGTTCTGGCG
MAACPSLPPCVLGCGGSG          50      ATGGCAGCATGCCCTTCTCTCCCTCCCTGCGTTCTTGGGTGTGGCGGTTCTGGCG
MAACPSLPPCFHRCGGSG          69      ATGGCAGCATGCCCCAGTCTTCCCCCCTGCTTTCACCGCTGTGGCGGTTCTGGCG
MAACPSLPPCTHRCGGSG          41      ATGGCAGCATGCCCGTCTCTTCCCCCCTGCACTCATAGGTGTGGCGGTTCTGGCG
MAACPTLPPCSYRCGGSG          40      ATGGCAGCATGCCCCACTTTGCCTCCTTGCTCTTACAGGTGTGGCGGTTCTGGCG

 Group8
MAACSNRCTLLPPCGGSG          104     ATGGCAGCATGCTCTAATCGGTGCACCTTGTTGCCGCCTTGTGGCGGTTCTGGCG
MAACSKRCNILPPCGGSG          49      ATGGCAGCATGCAGCAAGCGTTGCAATATTTTGCCCCCTTGTGGCGGTTCTGGCG
MAACTRRCLILPPCGGSG          42      ATGGCAGCATGCACTCGTCGCTGCCTTATTCTCCCTCCGTGTGGCGGTTCTGGCG
MAACNTKCSILPPCGGSG          103     ATGGCAGCATGCAACACTAAGTGCTCTATTCTCCCCCCCTGTGGCGGTTCTGGCG
MAACVSTCQILPPCGGSG          85      ATGGCAGCATGCGTTAGTACGTGCTAGATCCTCCCCCCTTGTGGCGGTTCTGGCG
MAACVMRCQVLPPCGGSG          65      ATGGCAGCATGCGTTATGCGTTGCTAGGTCCTGCCTCCCTGTGGCGGTTCTGGCG
MAACVDRCFILPPCGGSG          38      ATGGCAGCATGCGTTGATCGCTGCTTTATTCTTCCCCCGTGTGGCGGTTCTGGCG

 Group9
MAACSRHCLTLPPCGGSG          364     ATGGCAGCATGCAGTCGTCATTGCCTGACTCTTCCTCCGTGTGGCGGTTCTGGCG
MAACSKHCTTLPPCGGSG          37      ATGGCAGCATGCAGTAAGCATTGCACCACGCTCCCTCCTTGTGGCGGTTCTGGCG
MAACSRKCVELPPCGGSG          35      ATGGCAGCATGCTCCAGGAAGTGCGTGGAGTTGCCTCCGTGTGGCGGTTCTGGCG

 Group10
MAACKRTHCLPPCCGGSG          743     ATGGCAGCATGCAAGCGTACCCATTGCCTCCCCCCCTGTTGTGGCGGTTCTGGCG
MAACPRARCLPPCCGGSG          35      ATGGCAGCATGCCCTCGGGCTCGGTGCCTTCCTCCCTGCTGTGGCGGTTCTGGCG
MAACKSVCLPPCCGGSG           52      ATGGCAGCATGCAAGTCTGTCTGCCTTCCTCCGTGCTGTGGCGGTTCTGGCG
MAACRSITCLPPCCGGSG          43      ATGGCAGCATGCCGGTCGATTACCTGCCTTCCTCCGTGCTGTGGCGGTTCTGGCG
MAACRVACLPPCCGGSG           46      ATGGCAGCATGCCGTGTCGCCTGCCTTCCTCCTTGCTGTGGCGGTTCTGGCG
MAACRVMRCLPPCCGGSG          45      ATGGCAGCATGCCGTGTCATGCGGTGCCTTCCTCCGTGTTGTGGCGGTTCTGGCG

 Group11
MAACYQLPPCDHSCGGSG          140     ATGGCAGCATGCTACTAGTTGCCTCCCTGCGATCACAGTTGTGGCGGTTCTGGCG
MAACYLPPCDHSCGGSG           48      ATGGCAGCATGCTACCTGCCCCCGTGCGACCATTCCTGTGGCGGTTCTGGCG
MAACRELPPCGHSCGGSG          81      ATGGCAGCATGCCGTGAGTTGCCGCCTTGCGGTCATTCGTGTGGCGGTTCTGGCG

 Group12
MAACHSRCPTLPPCGGSG          91      ATGGCAGCATGCCACTCTCGCTGCCCCACTTTGCCCCCTTGTGGCGGTTCTGGCG
MAACHSRCPQLPPCGGSG          40      ATGGCAGCATGCCATAGTAGGTGCCCCTAGCTTCCCCCGTGTGGCGGTTCTGGCG
MAACDSRCPRLPPCGGSG          36      ATGGCAGCATGCGATAGTCGGTGCCCTCGGCTCCCCCCTTGTGGCGGTTCTGGCG
MAACTSRCPQLPPCGGSG          42      ATGGCAGCATGCACCTCGCGGTGCCCTCAGCTCCCTCCGTGTGGCGGTTCTGGCG
MAACTQRCPQLPPCGGSG          40      ATGGCAGCATGCACCTAGAGGTGCCCTCAGCTCCCCCCCTGTGGCGGTTCTGGCG

 Group13
MAACILPPCPSSCGGSG           68      ATGGCAGCATGCATTCTCCCCCCGTGCCCTTCTTCGTGTGGCGGTTCTGGCG
MAACILPPCPFSCGGSG           51      ATGGCAGCATGCATCTTGCCGCCTTGCCCCTTTTCCTGTGGCGGTTCTGGCG
MAACILPPCPYHCGGSG           102     ATGGCAGCATGCATCCTGCCCCCTTGCCCGTACCATTGTGGCGGTTCTGGCG
```

Group14		
MAACLLYSCEDQCGGSG	36	ATGGCAGCATGCCTCCTCTACAGTTGCGAGGACTAGTGTGGCGGTTCTGGCG
MAACMGYCSCWAQCGGSG	39	ATGGCAGCATGCATGGGTTATTGCTCGTGCTGGGCTTAGTGTGGCGGTTCTGGCG
MAACTILPPCPSYCGGSG	41	ATGGCAGCATGCACTATTTTGCCCCCTTGCCCCTCTTATTGTGGCGGTTCTGGCG
MAACHILPPCDSICGGSG	39	ATGGCAGCATGCCACATTTTGCCCCCTTGCGACTCTATTTGTGGCGGTTCTGGCG
MAACRILPPCADHCGGSG	45	ATGGCAGCATGCCGGATCCTCCCGCCTTGCGCTGATCATTGTGGCGGTTCTGGCG
MAACQLLPPCPFNCGGSG	78	ATGGCAGCATGCTAGCTCCTTCCGCCTTGCCCTTTTAATTGTGGCGGTTCTGGCG
MAACWVLPPCGSWCGGSG	55	ATGGCAGCATGCTGGGTGCTTCCTCCTTGCGGCTCTTGGTGTGGCGGTTCTGGCG
MAACPQLPPCKYICGGSG	36	ATGGCAGCATGCCCTTAGCTTCCCCCTTGCAAGTACATTTGTGGCGGTTCTGGCG
MAACGQLPPCSVVCGGSG	36	ATGGCAGCATGCGGGTAGCTCCCTCCCTGCTCCGTTGTCTGTGGCGGTTCTGGCG
MAACRQLPPCSFECGGSG	302	ATGGCAGCATGCAGGTAGCTTCCTCCTTGCTCTTTCGAGTGTGGCGGTTCTGGCG
MAACLQLPPCAWTCGGSG	64	ATGGCAGCATGCTTGCAGCTGCCCCCTTGCGCTTGGACGTGTGGCGGTTCTGGCG
MAACLQLPPCNVSCGGSG	96	ATGGCAGCATGCCTGTAGCTTCCTCCTTGCAACGTGTCCTGTGGCGGTTCTGGCG
MAACILPPCGLFYCGGSG	35	ATGGCAGCATGCATTCTCCCGCCCTGCGGTCTGTTTTATTGTGGCGGTTCTGGCG
MAACILPPCELNHCGGSG	39	ATGGCAGCATGCATTCTTCCTCCCTGCGAGCTTAATCATTGTGGCGGTTCTGGCG
MAACYLPPCSKSRCGGSG	57	ATGGCAGCATGCTACCTCCCTCCTTGCAGCAAGTCTCGGTGTGGCGGTTCTGGCG
MAACMLPPCFYDICGGSG	35	ATGGCAGCATGCATGCTCCCTCCTTGCTTCTATGATATTTGTGGCGGTTCTGGCG
Group15		
MAACLPPCPLLPCGGSG	54	ATGGCAGCATGCCTGCCTCCTTGCCCGTTGTTGCCGTGTGGCGGTTCTGGCG
MAACGPPCILPPCGGSG	72	ATGGCAGCATGCGGTCCCCCCTGCATTTTGCCTCCCTGTGGCGGTTCTGGCG
MAACFAPCNILPPCGGSG	43	ATGGCAGCATGCTTCGCTCCTTGCAACATCCTTCCCCCCTGTGGCGGTTCTGGCG
MAACFAPCPFLPPCGGSG	37	ATGGCAGCATGCTTCGCTCCTTGCCCGTTCTTGCCTCCTTGTGGCGGTTCTGGCG
MAACRNQCLILPPCGGSG	39	ATGGCAGCATGCCGCAACCAGTGCTTGATTCTTCCTCCCTGTGGCGGTTCTGGCG
MAACGGHCVILPPCGGSG	77	ATGGCAGCATGCGGTGGTCACTGCGTTATTCTCCCCCCGTGTGGCGGTTCTGGCG
MAACYGECQVLPPCGGSG	41	ATGGCAGCATGCTACGGCGAGTGCTAGGTGTTGCCTCCTTGTGGCGGTTCTGGCG
MAACYGQCTQLPPCGGSG	95	ATGGCAGCATGCTACGGTCAGTGCACTTAGCTGCCTCCTTGTGGCGGTTCTGGCG
MAACHGKCAFLPPCGGSG	36	ATGGCAGCATGCCACGGTAAGTGCGCGTTTTTGCCCCCCTGTGGCGGTTCTGGCG
MAACHSGCYVLPPCGGSG	36	ATGGCAGCATGCCACTCGGGGTGCTACGTTCTGCCTCCGTGTGGCGGTTCTGGCG
MAACGQVCLPPYCGGSG	81	ATGGCAGCATGCGGGTAGGTTTGCTTGCCTCCGTATTGTGGCGGTTCTGGCG
MAACYYTCLPPYCGGSG	49	ATGGCAGCATGCTACTATACGTGCTTGCCCCCGTATTGTGGCGGTTCTGGCG
MAACYALCLPPYQCGGSG	56	ATGGCAGCATGCTATGCGTTGTGCCTTCCCCCTTATCAGTGTGGCGGTTCTGGCG
Group16		
MAACSLICSIGSCGGSG	41	ATGGCAGCATGCTCCTTGATTTGCAGCATTGGTTCTTGTGGCGGTTCTGGCG
MAACSSICAIGLCGGSG	41	ATGGCAGCATGCAGTAGTATTTGCGCTATTGGGTTGTGTGGCGGTTCTGGCG
MAACNYQCTLAYCGGSG	45	ATGGCAGCATGCAACTACCAGTGCACTCTTGCGTATTGTGGCGGTTCTGGCG
MAACTTTCSFVYCGGSG	46	ATGGCAGCATGCACTACTACTTGCTCGTTTGTGTACTGTGGCGGTTCTGGCG
MAACMYDCGYTYCGGSG	37	ATGGCAGCATGCATGTATGACTGCGGCTATACCTACTGTGGCGGTTCTGGCG
MAACSYLCPQSFCGGSG	40	ATGGCAGCATGCTCTTATCTTTGCCCCTAGTCCTTTTGTGGCGGTTCTGGCG
MAACRAHCSVSLCGGSG	58	ATGGCAGCATGCCGTGCTCACTGCTCTGTCTCGCTCTGTGGCGGTTCTGGCG
MAACVRACQLSLCGGSG	62	ATGGCAGCATGCGTTCGTGCTTGCTAGCTCAGTCTTTGTGGCGGTTCTGGCG
MAACLRSNCSLVACGGSG	36	ATGGCAGCATGCCTCAGGAGTAATTGCTCGTTGGTTGCTTGTGGCGGTTCTGGCG
Group17		
MAACIQRSCVTHLCGGSG	37	ATGGCAGCATGCATTTAGCGTAGTTGCGTCACCCACCTTTGTGGCGGTTCTGGCG
MAACVQRCLLDLECGGSG	37	ATGGCAGCATGCGTTTAGCGTTGCCTGTTGGACCTTGAGTGTGGCGGTTCTGGCG
MAACTTNCVNQMCGGSG	37	ATGGCAGCATGCACCACTAATTGCGTGAACTAGATGTGTGGCGGTTCTGGCG
MAACPSRCMNVICGGSG	40	ATGGCAGCATGCCCTTCGAGGTGCATGAACGTGATCTGTGGCGGTTCTGGCG
Group18		
MAACNRTCHPQFPCGGSG	44	ATGGCAGCATGCAACCGGACGTGCCATCCCCAGTTCCCTTGTGGCGGTTCTGGCG
MAACEVGCHPQFACGGSG	44	ATGGCAGCATGCGAGGTGGGTTGCCATCCTTAGTTTGCTTGTGGCGGTTCTGGCG
MAACGIDCHPQGGCGGSG	93	ATGGCAGCATGCGGCATTGATTGCCATCCTTAGGGTGGGTGTGGCGGTTCTGGCG
Group MIXED		
MAACQRVCSRWQCGGSG	38	ATGGCAGCATGCTAGAGGGTTTGCAGTAGGTGGCAGTGTGGCGGTTCTGGCG
MAACSPICWRYKCGGSG	45	ATGGCAGCATGCTCTCCTATCTGCTGGCGGTACAAGTGTGGCGGTTCTGGCG
MAACHVNACYALSCGGSG	36	ATGGCAGCATGCCACGTTAACGCTTGCTATGCGTTGTCCTGTGGCGGTTCTGGCG
MAACHYTPCHQDSCGGSG	46	ATGGCAGCATGCCATTACACTCCTTGCCATTAGGATTCTTGTGGCGGTTCTGGCG
MAACNKGCNYPVCGGSG	35	ATGGCAGCATGCAACAAGGGTTGCAATTATCCTGTGTGTGGCGGTTCTGGCG
MAACLGGCYPVPCGGSG	41	ATGGCAGCATGCCTGGGCGGGTGCTATCCGGTTCCTTGTGGCGGTTCTGGCG
MAACHPQNCPLVFCGGSG	37	ATGGCAGCATGCCACCCTTAGAATTGCCCCCTCGTGTTTTGTGGCGGTTCTGGCG
MAACKGQTCPQVVCGGSG	51	ATGGCAGCATGCAAGGGGTAGACGTGCCCTTAGGTTGTGTGTGGCGGTTCTGGCG
MAACCNKRVCLDCCGGSG	39	ATGGCAGCATGCTGTAACAAGCGGGTGTGCCTTGACTGTTGTGGCGGTTCTGGCG
MAACAKRVCTSCCGGSG	43	ATGGCAGCATGCGCTAAGCGGGTGTGCACTTCCTGTTGTGGCGGTTCTGGCG
MAACSGYCRTGVSWGGSG	41	ATGGCAGCATGCTCGGGTTATTGCAGGACGGGGGTTTCTTGGGGCGGTTCTGGCG
MAARMKSSCLPPCCGGSG	106	ATGGCAGCACGCATGAAGAGTAGTTGCCTCCCTCCGTGCTGTGGCGGTTCTGGCG
MAACKVGCLYSWCGGSG	47	ATGGCAGCATGCAAGGTTGGCTGCTTGTATAGTTGGTGTGGCGGTTCTGGCG
MAACQQSCLYKACGGSG	35	ATGGCAGCATGCTAGCAGTCGTGCTTGTACAAGGCGTGTGGCGGTTCTGGCG
MAACYRACSFKLVCGGSG	65	ATGGCAGCATGCTATCGTGCGTGCAGTTTTAAGTTGGTGTGGCGGTTCTGGCG
MAACQVFCSFDMFCGGSG	41	ATGGCAGCATGCTAGGTTTTCTGCTCGTTCGATATGTTTTGTGGCGGTTCTGGCG
MAACERVCHVDSACGGSG	39	ATGGCAGCATGCGAGCGTGTGTGCCATGTGGATTCTGCGTGTGGCGGTTCTGGCG
MAACQSICAVPGFCGGSG	35	ATGGCAGCATGCTAGTCGATCTGCGCTGTCCCTGGCTTTTGTGGCGGTTCTGGCG
MAACFGSCFNLLYCGGSG	41	ATGGCAGCATGCTTCGGTAGTTGCTTTAATCTCCTGTATTGTGGCGGTTCTGGCG
MAACLHSCSDECLCGGSG	45	ATGGCAGCATGCTTGCACAGCTGCAGCGATGAGTGTCTTTGTGGCGGTTCTGGCG
MAACQHRCCDVHFCGGSG	36	ATGGCAGCATGCTAGCATCGTTGTTGCGATGTGCATTTTTGTGGCGGTTCTGGCG
MAACCSLSCDLQTCGGSG	41	ATGGCAGCATGCTGCAGTCTTTCCTGCGACCTCTAGACTTGTGGCGGTTCTGGCG

MAACDALMCTHSSCGGSG	36	ATGGCAGCATGCGATGCTTTGATGTGCACCCATAGCTCTTGTGGCGGTTCTGGCG
MAACSSSICNHICGGSG	37	ATGGCAGCATGCTCCAGTAGTATTTGCAATCACATTTGTGGCGGTTCTGGCG
MAACFGSSCCVICGGSG	72	ATGGCAGCATGCTTTGGGTCTAGCTGCTGTGTTATCTGTGGCGGTTCTGGCG
MAACNSSVCSCIICGGSG	46	ATGGCAGCATGCAATTCTAGTGTTTGCTCGTGTATCATTTGTGGCGGTTCTGGCG
MAACPVVCQWFTCGGSG	64	ATGGCAGCATGCCCTGTCGTTTGCTAGTGGTTCACCTGTGGCGGTTCTGGCG
MAACRFTCVKVPCGGSG	42	ATGGCAGCATGCAGGTTTACCTGCGTGAAGGTTCCCTGTGGCGGTTCTGGCG
MAACDYYTCWQFPCGGSG	36	ATGGCAGCATGCGACTATTATACGTGCTGGTAGTTCCCTTGTGGCGGTTCTGGCG
MAACCYACFQPCGCGGSG	41	ATGGCAGCATGCTGTTACGCCTGCTTTTAGCCTTGTGGTTGTGGCGGTTCTGGCG
MAACQELVLPPCFCGGSG	74	ATGGCAGCATGCTAGGAGTTGGTGCTCCCGCCATGTTTTTGTGGCGGTTCTGGCG
MAACLTYCLPTGCGGSG	44	ATGGCAGCATGCCTCACCTACTGCCTTCCGACTGGGTGTGGCGGTTCTGGCG
MAACMNICLPPTNCGGSG	43	ATGGCAGCATGCATGAATATTTGCTTGCCTCCGACCAATTGTGGCGGTTCTGGCG
MAACSQFCLPPFSCGGSG	46	ATGGCAGCATGCAGTCAGTTTTGCTTGCCTCCGTTCTCTTGTGGCGGTTCTGGCG
MAACSICCMSEIQCGGSG	62	ATGGCAGCATGCAGCATCTGTTGCATGTCTGAGATTTAGTGTGGCGGTTCTGGCG
MAACPAGCMTTCSCGGSG	47	ATGGCAGCATGCCCTGCGGGTTGCATGACTACCTGTTCGTGTGGCGGTTCTGGCG
MAACQGTCLLCGDCGGSG	42	ATGGCAGCATGCCAGGGTACTTGCCTGTTGTGTGGTGACTGTGGCGGTTCTGGCG
MAACLRYCTQLPPCGGSG	40	ATGGCAGCATGCTTGCGTTATTGCACCTAGCTTCCCCCCTGTGGCGGTTCTGGCG
MAACPSYCTQGLPCGGSG	42	ATGGCAGCATGCCCTTCTTATTGCACTTAGGGCCTTCCCTGTGGCGGTTCTGGCG
MAACIVQWCRVPCGGSG	45	ATGGCAGCATGCATTGTTTAGTGGTGCCGTGTTCCGTGTGGCGGTTCTGGCG
MAACLPPCYHHFLCGGSG	37	ATGGCAGCATGCCTTCCCCCGTGCTATCATCACTTTCTCTGTGGCGGTTCTGGCG
MAACLPPYSWTDCGGSG	36	ATGGCAGCATGCCTGCCCCCTTACAGCTGGACCGACTGTGGCGGTTCTGGCG
MAACSRYNMCPFMCGGSG	35	ATGGCAGCATGCTCCCGGTATAATATGTGCCCGTTCATGTGTGGCGGTTCTGGCG
MAACNLYPLCSPYCGGSG	35	ATGGCAGCATGCAATCTGTATCCTCTCTGCAGCCCCTACTGTGGCGGTTCTGGCG
MAACGWFYCSWTCGGSG	78	ATGGCAGCATGCGGTTGGTTCTATTGCTCTTGGACGTGTGGCGGTTCTGGCG

SrtA – Library B

Group1		
MAACAILPPCGQLSCGGSG	290	ATGGCAGCATGCGCTATTCTGCCGCCGTGCGGGCAGCTTAGTTGTGGCGGTTCTGGCG
MAACAYLPPCGSMLCGGSG	67	ATGGCAGCATGCGCTTATCTTCCTCCGTGCGGTTCTATGTTGTGTGGCGGTTCTGGCG
MAACQLLPPCQFLQCGGSG	311	ATGGCAGCATGCCAGCTTCTGCCTCCTTGCTAGTTTTTGCAGTGTGGCGGTTCTGGCG
MAACQYLPPCSFLVCGGSG	65	ATGGCAGCATGCCAGTATCTTCCTCCTTGCTCTTTTTTGGTTTGTGGCGGTTCTGGCG
MAACSVLPPCSFVACGGSG	773	ATGGCAGCATGCTCGGTGTTGCCTCCTTGCTCTTTTGTTGCGTGTGGCGGTTCTGGCG
MAACSLLPPCTFVACGGSG	233	ATGGCAGCATGCTCGTTGCTGCCGCCGTGCACTTTTGTGGCGTGTGGCGGTTCTGGCG
MAACALLPPCSWVSCGGSG	229	ATGGCAGCATGCGCGTTGCTGCCGCCGTGCTCGTGGGTGTCTTGTGGCGGTTCTGGCG
MAACAILPPCHFRSCGGSG	38	ATGGCAGCATGCGCGATTCTTCCGCCTTGCCATTTTCGGTCGTGTGGCGGTTCTGGCG
MAACSRLPPCVILDCGGSG	237	ATGGCAGCATGCTCTAGGCTTCCTCCTTGCGTTATTCTTGATTGTGGCGGTTCTGGCG
MAACTRLPPCVQLSCGGSG	35	ATGGCAGCATGCACTCGTTTGCCGCCTTGCGTTTAGCTTTCGTGTGGCGGTTCTGGCG
MAACSQLPPCTYLSCGGSG	283	ATGGCAGCATGCTCTCAGCTTCCGCCTTGCACTTATCTTTCGTGTGGCGGTTCTGGCG
MAACSVLPPCIQWSCGGSG	198	ATGGCAGCATGCTCTGTTCTGCCTCCGTGCATTCAGTGGAGTTGTGGCGGTTCTGGCG
MAACELLPPCLLSECGGSG	166	ATGGCAGCATGCGAGCTTTTGCCGCCTTGCCTGCTGTCGGAGTGTGGCGGTTCTGGCG
MAACALLPPCFIQECGGSG	69	ATGGCAGCATGCGCTCTGTTGCCGCCGTGCTTTATTTAGGAGTGTGGCGGTTCTGGCG
MAACLTLPPCVLQQCGGSG	147	ATGGCAGCATGCTTGACGCTGCCGCCGTGCGTGCTGCAGTAGTGTGGCGGTTCTGGCG
MAACTLLPPCVFQQCGGSG	33	ATGGCAGCATGCACGCTGTTGCCTCCGTGCGTTTTTCAGCAGTGTGGCGGTTCTGGCG
MAACLVLPPCFLVDCGGSG	789	ATGGCAGCATGCCTTGTTCTTCCGCCTTGCTTTTTGGTGGATTGTGGCGGTTCTGGCG
MAACLVLPPCMIMECGGSG	131	ATGGCAGCATGCCTTGTGCTTCCTCCGTGCATGATTATGGAGTGTGGCGGTTCTGGCG
MAACQVLPPCGLQLCGGSG	2302	ATGGCAGCATGCCAGGTGTTGCCTCCGTGCGGTCTGTAGCTGTGTGGCGGTTCTGGCG
MAACQLLPPCAIQWCGGSG	661	ATGGCAGCATGCCAGTTGCTTCCGCCGTGCGCGATTTAGTGGTGTGGCGGTTCTGGCG
MAACQLLPPCVDKNCGGSG	65	ATGGCAGCATGCTAGCTTTTGCCTCCTTGCGTGGATAAGAATTGTGGCGGTTCTGGCG
Group2		
MAACLAKRCLTLPPCGGSG	227	ATGGCAGCATGCCTTGCTAAGCGTTGCCTGACTCTTCCTCCTTGTGGCGGTTCTGGCG
MAACAGNRCLLLPPCGGSG	65	ATGGCAGCATGCGCGGGTAATCGGTGCTTGCTTCTTCCTCCTTGTGGCGGTTCTGGCG
MAACRGRTCLVLPPCGGSG	78	ATGGCAGCATGCCGGGGGCGTACTTGCTTGGTGCTTCCTCCGTGTGGCGGTTCTGGCG
MAACVAFTCPILPPCGGSG	439	ATGGCAGCATGCGTGGCGTTTACGTGCCCGATTCTGCCTCCGTGTGGCGGTTCTGGCG
MAACIGSLCPVLPPCGGSG	86	ATGGCAGCATGCATTGGTAGTTTGTGCCCGGTTCTTCCTCCGTGTGGCGGTTCTGGCG
MAACLLASCPILPPCGGSG	316	ATGGCAGCATGCTTGCTGGCTTCGTGCCCGATTCTTCCTCCGTGTGGCGGTTCTGGCG
MAACRMSSCPILPPCGGSG	52	ATGGCAGCATGCCGTATGTCTTCGTGCCCTATTTTGCCTCCTTGTGGCGGTTCTGGCG
MAACQSTFCPILPPCGGSG	492	ATGGCAGCATGCTAGTCTACGTTTTGCCCGATTCTGCCTCCTTGTGGCGGTTCTGGCG
MAACEGLMCPLLPPCGGSG	33	ATGGCAGCATGCGAGGGTCTGATGTGCCCTCTTCTTCCTCCGTGTGGCGGTTCTGGCG
MAACPNQTCPLLPPCGGSG	209	ATGGCAGCATGCCCTAATTAGACGTGCCCTTTGCTGCCTCCTTGTGGCGGTTCTGGCG
MAACDAWRCPVLPPCGGSG	61	ATGGCAGCATGCGATGCTTGGAGGTGCCCTGTGTTGCCTCCGTGTGGCGGTTCTGGCG
MAACVFSECSILPPCGGSG	134	ATGGCAGCATGCGTTTTTTCTGAGTGCTCTATTCTTCCTCCGTGTGGCGGTTCTGGCG
MAACTLSFCSVLPPCGGSG	34	ATGGCAGCATGCACGCTGTCTTTTTGCAGTGTTCTTCCTCCTTGTGGCGGTTCTGGCG
MAACVYQYCAILPPCGGSG	44	ATGGCAGCATGCGTGTATCAGTATTGCGCGATTTTGCCGCCTTGTGGCGGTTCTGGCG
MAACQYLHCSVLPPCGGSG	138	ATGGCAGCATGCTAGTATTTGCATTGCTCTGTTCTGCCTCCGTGTGGCGGTTCTGGCG
MAACHYTSCVILPPCGGSG	57	ATGGCAGCATGCCATTATACGAGTTGCGTTATTTTGCCGCCTTGTGGCGGTTCTGGCG
MAACNESSCTILPPCGGSG	42	ATGGCAGCATGCAATGAGTCTTCGTGCACTATTCTGCCTCCTTGTGGCGGTTCTGGCG
MAACQLGSCMVLPPCGGSG	61	ATGGCAGCATGCTAGCTTGGTTCTTGCATGGTGTTGCCTCCGTGTGGCGGTTCTGGCG
Group3		
MAACILPPCPTSEYCGGSG	82	ATGGCAGCATGCATTCTGCCTCCTTGCCCGACTTCGGAGTATTGTGGCGGTTCTGGCG
MAACILPPCPTTIPCGGSG	49	ATGGCAGCATGCATTCTTCCGCCGTGCCCGACGACGATTCCGTGTGGCGGTTCTGGCG
MAACYLPPCPSLPHCGGSG	66	ATGGCAGCATGCTATCTTCCTCCGTGCCCGTCTCTGCCTCATTGTGGCGGTTCTGGCG
Group4		
MAACPQKQCVWLPPCGGSG	63	ATGGCAGCATGCCCTCAGAAGCAGTGCGTTTGGCTTCCTCCTTGTGGCGGTTCTGGCG

MAACGAKYCSYLPPCGGSG	78	ATGGCAGCATGCGGGGCGAAGTATTGCTCGTATTTGCCTCCTTGTGGCGGTTCTGGCG
MAACCARLCAYLPPCGGSG	59	ATGGCAGCATGCTGTGCGCGTCTGTGCGCTTATTTGCCGCCTTGTGGCGGTTCTGGCG
Group5		
MAACRQLPPCAEYVCGGSG	1231	ATGGCAGCATGCCGGTAGCTGCCTCCTTGCGCTGAGTATGTTTGTGGCGGTTCTGGCG
MAACTLLPPCAEYVCGGSG	33	ATGGCAGCATGCACGCTGTTGCCTCCTTGCGCTGAGTATGTTTGTGGCGGTTCTGGCG
MAACRVLPPCSEYNCGGSG	197	ATGGCAGCATGCAGGGTGTTGCCTCCGTGCTCTGAGTATAATTGTGGCGGTTCTGGCG
MAACTRLPPCSEFFCGGSG	109	ATGGCAGCATGCACGCGGCTTCCGCCTTGCTCTGAGTTTTTTTGTGGCGGTTCTGGCG
MAACRTLPPCSDLTCGGSG	146	ATGGCAGCATGCCGTACGCTGCCTCCTTGCTCTGATCTTACTTGTGGCGGTTCTGGCG
MAACSLLPPCGSTTCGGSG	102	ATGGCAGCATGCTCTTTGTTGCCTCCTTGCGGTTCTACTACGTGTGGCGGTTCTGGCG
MAACSLLPPCHSTNCGGSG	66	ATGGCAGCATGCAGTCTTCTTCCTCCGTGCCATTCGACGAATTGTGGCGGTTCTGGCG
MAACTLLPPCSSLQCGGSG	325	ATGGCAGCATGCACTTTGTTGCCGCCGTGCTCTTCTCTTCAGTGTGGCGGTTCTGGCG
MAACRLLPPCSSSICGGSG	40	ATGGCAGCATGCCGTCTGTTGCCTCCTTGCTCGTCGTCTATTTGTGGCGGTTCTGGCG
MAACSSLPPCSTSECGGSG	53	ATGGCAGCATGCTCGAGTTTGCCTCCTTGCTCGACTTCTGAGTGTGGCGGTTCTGGCG
MAACQILPPCHSPGCGGSG	410	ATGGCAGCATGCTAGATTCTGCCTCCTTGCCATTCGCCGGGGTGTGGCGGTTCTGGCG
MAACSVLPPCKSQGCGGSG	61	ATGGCAGCATGCTCTGTGTTGCCGCCGTGCAAGAGTTAGGGGTGTGGCGGTTCTGGCG
MAACSLLPPCTPVLCGGSG	130	ATGGCAGCATGCTCTCTGCTTCCGCCTTGCACTCCTGTTTTGTGTGGCGGTTCTGGCG
MAACSLLPPCSPILCGGSG	64	ATGGCAGCATGCAGTTTGCTGCCTCCTTGCTCACCGATTCTTTGTGGCGGTTCTGGCG
MAACLLLPPCAPTTCGGSG	282	ATGGCAGCATGCTTGTTGCTGCCGCCTTGCGCGCCGACTACTTGTGGCGGTTCTGGCG
MAACSILPPCNPPQCGGSG	5860	ATGGCAGCATGCTCGATTCTTCCTCCGTGCAATCCTCCGTAGTGTGGCGGTTCTGGCG
MAACTLLPPCTPDQCGGSG	3990	ATGGCAGCATGCACGCTGTTGCCTCCTTGCACGCCGGATTAGTGTGGCGGTTCTGGCG
MAACPILPPCPPTQCGGSG	165	ATGGCAGCATGCCCTATTTTGCCTCCTTGCCCGCCGACTCAGTGTGGCGGTTCTGGCG
MAACNLLPPCPPLSCGGSG	114	ATGGCAGCATGCAATCTGCTGCCTCCTTGCCCGCCTCTGTCTTGTGGCGGTTCTGGCG
MAACLLLPPCPIMTCGGSG	340	ATGGCAGCATGCCTTCTGCTTCCGCCTTGCCCGATTATGACGTGTGGCGGTTCTGGCG
MAACVVLPPCPLQTCGGSG	58	ATGGCAGCATGCGTGGTGTTGCCTCCGTGCCCGTTGCAGACTTGTGGCGGTTCTGGCG
MAACLQLPPCPLYSCGGSG	71	ATGGCAGCATGCCTTTAGCTTCCGCCTTGCCCGTTGTATTCTTGTGGCGGTTCTGGCG
MAACLTLPPCPSFTCGGSG	295	ATGGCAGCATGCTTGACTTTGCCGCCGTGCCCGTCTTTTACTTGTGGCGGTTCTGGCG
MAACIVLPPCNSFTCGGSG	65	ATGGCAGCATGCATTGTGTTACCTCCTTGCAATTCTTTTACGTGTGGCGGTTCTGGCG
MAACYLLPPCPSTSCGGSG	204	ATGGCAGCATGCTATTTGTTGCCGCCGTGCCCTTCGACGTCTTGTGGCGGTTCTGGCG
MAACLLLPPCPSINCGGSG	170	ATGGCAGCATGCCTTCTGCTTCCTCCGTGCCCTTCGATTAATTGTGGCGGTTCTGGCG
MAACKLLPPCPKVVCGGSG	43	ATGGCAGCATGCAAGCTTCTTCCTCCTTGCCCTAAGGTTGTGTGTGGCGGTTCTGGCG
MAACMVLPPCPHQRCGGSG	222	ATGGCAGCATGCATGGTGTTGCCTCCGTGCCCTCATCAGCGGTGTGGCGGTTCTGGCG
MAACYELPPCPQLFCGGSG	36	ATGGCAGCATGCTATGAGTTGCCGCCGTGCCCTCAGCTGTTTTGTGGCGGTTCTGGCG
MAACLTLPPCQAVSCGGSG	419	ATGGCAGCATGCCTTACTCTTCCTCCTTGCTAGGCGGTTTCTTGTGGCGGTTCTGGCG
MAACLLLPPCQRTSCGGSG	34	ATGGCAGCATGCCTTTTGTTGCCTCCGTGCTAGAGGACTTCTTGTGGCGGTTCTGGCG
MAACIQLPPCQQSSCGGSG	279	ATGGCAGCATGCATTCAGCTTCCGCCTTGCTAGCAGTCTTCGTGTGGCGGTTCTGGCG
MAACLELPPCQFASCGGSG	625	ATGGCAGCATGCTTGGAGTTGCCGCCGTGCCAGTTTGCTTCGTGTGGCGGTTCTGGCG
MAACLTLPPCATTYCGGSG	190	ATGGCAGCATGCTTGACGTTGCCTCCGTGCGCTACGACTTATTGTGGCGGTTCTGGCG
MAACLTLPPCLSTLCGGSG	99	ATGGCAGCATGCTTGACTCTGCCTCCTTGCCTGTCTACGCTTTGTGGCGGTTCTGGCG
MAACYLLPPCAISVCGGSG	243	ATGGCAGCATGCTATCTTTTGCCGCCTTGCGCGATTTCGGTTTGTGGCGGTTCTGGCG
MAACFLLPPCQVALCGGSG	66	ATGGCAGCATGCTTTCTTTTGCCTCCTTGCTAGGTTGCGCTTTGTGGCGGTTCTGGCG
MAACMLLPPCQLNFCGGSG	87	ATGGCAGCATGCATGTTGCTGCCTCCGTGCCAGCTGAATTTTTGTGGCGGTTCTGGCG
MAACLQLPPCQVIVCGGSG	285	ATGGCAGCATGCCTGCAGCTGCCTCCTTGCTAGGTTATTGTTTGTGGCGGTTCTGGCG
MAACLQLPPCGVSLCGGSG	139	ATGGCAGCATGCCTGTAGCTGCCGCCTTGCGGTGTTAGTCTGTGTGGCGGTTCTGGCG
MAACIMLPPCSVIRCGGSG	68	ATGGCAGCATGCATTATGCTTCCGCCTTGCTCGGTTATTCGTTGTGGCGGTTCTGGCG
MAACLLLPPCSIQGCGGSG	38	ATGGCAGCATGCCTTTTGTTGCCGCCTTGCTCTATTTAGGGTTGTGGCGGTTCTGGCG
Group6		
MAACEILPPCLQFQCGGSG	607	ATGGCAGCATGCGAGATTTTGCCTCCTTGCCTTTAGTTTCAGTGTGGCGGTTCTGGCG
MAACQELPPCLQIQCGGSG	53	ATGGCAGCATGCCAGGAGCTTCCGCCTTGCCTGCAGATTTAGTGTGGCGGTTCTGGCG
MAACWLLPPCLQMQCGGSG	127	ATGGCAGCATGCTGGCTTCTGCCTCCGTGCCTGCAGATGTAGTGTGGCGGTTCTGGCG
MAACWLLPPCLTILCGGSG	49	ATGGCAGCATGCTGGCTTCTTCCTCCGTGCTTGACGATTCTGTGTGGCGGTTCTGGCG
MAACGLLPPCHQFHCGGSG	206	ATGGCAGCATGCGGGCTTTTGCCGCCTTGCCATTAGTTTCATTGTGGCGGTTCTGGCG
MAACPSLPPCWQLQCGGSG	1131	ATGGCAGCATGCCCTAGTTTGCCTCCTTGCTGGCAGTTGTAGTGTGGCGGTTCTGGCG
MAACSSLPPCYPFQCGGSG	35	ATGGCAGCATGCAGTTCGTTGCCTCCTTGCTATCCGTTTTAGTGTGGCGGTTCTGGCG
Group7		
MAACGYRTCYSLPPCGGSG	76	ATGGCAGCATGCGGGTATCGTACTTGCTATTCTCTGCCTCCTTGTGGCGGTTCTGGCG
MAACGLQRCYLLPPCGGSG	73	ATGGCAGCATGCGGTTTGTAGAGGTGCTATCTTTTGCCTCCTTGTGGCGGTTCTGGCG
MAACRFQFCHQLPPCGGSG	39	ATGGCAGCATGCAGGTTTCAGTTTTGCCATCAGTTGCCTCCGTGTGGCGGTTCTGGCG
Group8		
MAACPMLPPCDLSYCGGSG	1258	ATGGCAGCATGCCCTATGCTGCCTCCGTGCGATCTGAGTTATTGTGGCGGTTCTGGCG
MAACPLLPPCHLSFCGGSG	537	ATGGCAGCATGCCCGTTGCTGCCTCCGTGCCATCTTTCTTTTTGTGGCGGTTCTGGCG
MAACPLLPPCHLPQCGGSG	4327	ATGGCAGCATGCCCGTTGCTTCCGCCGTGCCATCTTCCTTAGTGTGGCGGTTCTGGCG
MAACPGLPPCMFSQCGGSG	45	ATGGCAGCATGCCCTGGTTTGCCGCCTTGCATGTTTTCTTAGTGTGGCGGTTCTGGCG
MAACPTLPPCQPGYCGGSG	329	ATGGCAGCATGCCCGACGTTGCCTCCTTGCCAGCCTGGTTATTGTGGCGGTTCTGGCG
MAACPSLPPCLPTICGGSG	199	ATGGCAGCATGCCCGAGTTTGCCGCCGTGCTTGCCGACGATTTGTGGCGGTTCTGGCG
MAACPSLPPCFNTYCGGSG	230	ATGGCAGCATGCCCGTCGCTTCCGCCTTGCTTTAATACTTATTGTGGCGGTTCTGGCG
MAACPQLPPCIHSYCGGSG	38	ATGGCAGCATGCCCTCAGCTGCCTCCGTGCATTCATTCTTATTGTGGCGGTTCTGGCG
MAACPLLPPCRNTVCGGSG	393	ATGGCAGCATGCCCTCTGTTGCCTCCTTGCCGGAATACGGTTTGTGGCGGTTCTGGCG
MAACPTLPPCHNQLCGGSG	114	ATGGCAGCATGCCCGACGCTTCCTCCGTGCCATAATTAGTTGTGTGGCGGTTCTGGCG
MAACPVLPPCFTRLCGGSG	35	ATGGCAGCATGCCCGGTTCTGCCTCCGTGCTTTACTCGTTTGTGTGGCGGTTCTGGCG
MAACPVLPPCRTIWCGGSG	467	ATGGCAGCATGCCCGGTGCTGCCTCCGTGCCGTACTATTTGGTGTGGCGGTTCTGGCG
MAACPVLPPCTRVHCGGSG	303	ATGGCAGCATGCCCTGTTCTGCCTCCTTGCACGCGGGTTCATTGTGGCGGTTCTGGCG
Group9		
MAACFWLDCTILPPCGGSG	85	ATGGCAGCATGCTTTTGGCTGGATTGCACGATTCTTCCGCCGTGTGGCGGTTCTGGCG

```
MAACTWLQCTILPPCGGSG          64      ATGGCAGCATGCACGTGGTTGTAGTGCACGATTCTGCCTCCTTGTGGCGGTTCTGGCG
MAACVWNNCSLLPPCGGSG          723     ATGGCAGCATGCGTGTGGAATAATTGCTCTCTTCTGCCTCCTTGTGGCGGTTCTGGCG

 Group10
MAACLLLPPCQHQMCGGSG          75      ATGGCAGCATGCCTTCTGCTGCCGCCTTGCCAGCATTAGATGTGTGGCGGTTCTGGCG
MAACTLLPPCLNQLCGGSG          37      ATGGCAGCATGCACGTTGCTGCCTCCGTGCCTGAATTAGCTGTGTGGCGGTTCTGGCG
MAACTVLPPCEQALCGGSG          38      ATGGCAGCATGCACTGTGTTGCCTCCTTGCGAGCAGGCTCTTTGTGGCGGTTCTGGCG
MAACYVLPPCSNVVCGGSG          188     ATGGCAGCATGCTATGTTCTTCCTCCGTGCTCTAATGTGGTGTGTGGCGGTTCTGGCG
MAACYILPPCSREMCGGSG          60      ATGGCAGCATGCTATATTTTGCCTCCGTGCTCTCGTGAGATGTGTGGCGGTTCTGGCG
MAACLILPPCSDLLCGGSG          352     ATGGCAGCATGCCTTATTCTGCCGCCTTGCTCGGATCTTCTTTGTGGCGGTTCTGGCG
MAACVILPPCSRLLCGGSG          105     ATGGCAGCATGCGTTATTCTTCCGCCTTGCTCGCGGTTGTTGTGTGGCGGTTCTGGCG

 Group11
MAACPELPPCQLMLCGGSG          2326    ATGGCAGCATGCCCGGAGCTTCCGCCGTGCTAGTTGATGCTGTGTGGCGGTTCTGGCG
MAACPELPPCLLLFCGGSG          35      ATGGCAGCATGCCCTGAGTTGCCTCCGTGCTTGTTGTTGTTTTGTGGCGGTTCTGGCG
MAACPELPPCTVLKCGGSG          554     ATGGCAGCATGCCCTGAGCTGCCGCCTTGCACGGTTCTTAAATGTGGCGGTTCTGGCG

 Group12
MAACAQLPPCDYSGCGGSG          505     ATGGCAGCATGCGCGTAGCTTCCTCCGTGCGATTATTCTGGGTGTGGCGGTTCTGGCG
MAACDILPPCVYTKCGGSG          130     ATGGCAGCATGCGATATTCTTCCTCCTTGCGTGTATACTAAGTGTGGCGGTTCTGGCG
MAACTELPPCGWLKCGGSG          115     ATGGCAGCATGCACTGAGCTTCCTCCTTGCGGTTGGCTTAAGTGTGGCGGTTCTGGCG
MAACRYLPPCPYKLCGGSG          616     ATGGCAGCATGCCGGTATCTTCCTCCTTGCCCTTATAAGCTGTGTGGCGGTTCTGGCG
MAACPYLPPCSWDLCGGSG          576     ATGGCAGCATGCCCTTATCTTCCTCCGTGCTCGTGGGATCTGTGTGGCGGTTCTGGCG
MAACQQLPPCMPRFCGGSG          54      ATGGCAGCATGCCAGTAGTTGCCTCCTTGCATGCCTCGGTTTTGTGGCGGTTCTGGCG
MAACTQLPPCTPRRCGGSG          36      ATGGCAGCATGCACGTAGCTTCCTCCTTGCACTCCTAGGCGTTGTGGCGGTTCTGGCG
MAACHYLPPCQPAICGGSG          88      ATGGCAGCATGCCATTATTTGCCGCCTTGCTAGCCGGCGATTTGTGGCGGTTCTGGCG
MAACPMLPPCGIFPCGGSG          120     ATGGCAGCATGCCCGATGCTTCCGCCTTGCGGGATTTTTCCTTGTGGCGGTTCTGGCG
MAACRVLPPCQVSPCGGSG          51      ATGGCAGCATGCCGTGTTCTTCCGCCTTGCTAGGTGTCTCCGTGTGGCGGTTCTGGCG
MAACLSLPPCGRPACGGSG          36      ATGGCAGCATGCTTGTCGTTGCCTCCGTGCGGTAGGCCTGCGTGTGGCGGTTCTGGCG
MAACHLLPPCGRQSCGGSG          250     ATGGCAGCATGCCATCTTTTGCCTCCGTGCGGTCGTTAGTCGTGTGGCGGTTCTGGCG
MASCPMLPPCMKHSCGGSG          37      ATGGCATCATGCCCTATGCTTCCTCCTTGCATGAAGCATTCGTGTGGCGGTTCTGGCG

 Group13
MAACMQVWCHPQGGCGGSG          61      ATGGCAGCATGCATGCAGGTGTGGTGCCATCCTCAGGGTGGGTGTGGCGGTTCTGGCG
MAACRWWVCHPQSGCGGSG          66      ATGGCAGCATGCCGGTGGTGGGTTTGCCATCCGTAGTCGGGTTGTGGCGGTTCTGGCG
MAACRFAYCHPQGDCGGSG          153     ATGGCAGCATGCCGGTTTGCGTATTGCCATCCTTAGGGGGATTGTGGCGGTTCTGGCG
MAACAYSSCHPQAPCGGSG          507     ATGGCAGCATGCGCTTATAGTTCTTGCCATCCTCAGGCTCCGTGTGGCGGTTCTGGCG
MAACPVTECHPQVFCGGSG          133     ATGGCAGCATGCCCTGTGACTGAGTGCCATCCTTAGGTGTTTTGTGGCGGTTCTGGCG

 Group14
MAACLLPPCSWWYDCGGSG          74      ATGGCAGCATGCCTGCTTCCTCCTTGCTCGTGGTGGTATGATTGTGGCGGTTCTGGCG
MAACVLPPCHWSLQCGGSG          133     ATGGCAGCATGCGTTCTTCCTCCTTGCCATTGGTCGCTGCAGTGTGGCGGTTCTGGCG
MAACILPPCDFRFACGGSG          113     ATGGCAGCATGCATTCTTCCGCCTTGCGATTTTCGTTTTGCGTGTGGCGGTTCTGGCG
MAACLLPPCGYRWGCGGSG          41      ATGGCAGCATGCTTGTTGCCGCCGTGCGGTTATCGTTGGGGTTGTGGCGGTTCTGGCG
MAACELLPCIFRWMCGGSG          46      ATGGCAGCATGCGAGTTGTTGCCGTGCATTTTTCGTTGGATGTGTGGCGGTTCTGGCG

 Group15
MAACLAKCSQLSRHCGGSG          34      ATGGCAGCATGCCTTGCGAAGTGCTCGTAGCTGTCTCGTCATTGTGGCGGTTCTGGCG
MAACLDRCNLLPPYCGGSG          36      ATGGCAGCATGCCTTGATCGGTGCAATTTGCTTCCTCCTTATTGTGGCGGTTCTGGCG
MAACQLGCSLLRPMCGGSG          56      ATGGCAGCATGCTAGCTTGGGTGCTCGCTGTTGCGTCCGATGTGTGGCGGTTCTGGCG

 Group16
MAACPTKSCLLLPPCGGSG          156     ATGGCAGCATGCCCTACGAAGTCTTGCCTTTTGCTTCCTCCGTGTGGCGGTTCTGGCG
MAACSSKFCLLLPPCGGSG          130     ATGGCAGCATGCTCTTCTAAGTTTTGCTTGTTGCTTCCTCCTTGTGGCGGTTCTGGCG
MAACTSNSCQFTPPCGGSG          36      ATGGCAGCATGCACTTCTAATTCGTGCTAGTTTACTCCTCCTTGTGGCGGTTCTGGCG
MAACASGTCTTLPPCGGSG          43      ATGGCAGCATGCGCGTCTGGTACGTGCACTACTCTTCCTCCGTGTGGCGGTTCTGGCG
MAACQSPTCHTLPPCGGSG          34      ATGGCAGCATGCTAGTCGCCGACTTGCCATACTTTGCCTCCGTGTGGCGGTTCTGGCG
MAACVPLRCTLLPPCGGSG          124     ATGGCAGCATGCGTTCCTCTTCGGTGCACTCTTCTGCCTCCGTGTGGCGGTTCTGGCG
MAACVPYPCNLLPPCGGSG          55      ATGGCAGCATGCGTTCCTTATCCGTGCAATTTGTTGCCGCCTTGTGGCGGTTCTGGCG
MAACLTIPCGLLPPCGGSG          39      ATGGCAGCATGCCTTACTATTCCTTGCGGTCTTCTTCCTCCGTGTGGCGGTTCTGGCG
MAACPRRCCTFLPPCGGSG          87      ATGGCAGCATGCCCGAGGCGTTGTTGCACGTTTCTGCCTCCGTGTGGCGGTTCTGGCG
MAACIQRTCLFLPPCGGSG          139     ATGGCAGCATGCATTCAGCGGACTTGCTTGTTTCTTCCGCCGTGTGGCGGTTCTGGCG
MAACMNSQCIELPPCGGSG          34      ATGGCAGCATGCATGAATTCGTAGTGCATTGAGTTGCCTCCTTGTGGCGGTTCTGGCG
MAACFTAKCLQLPPCGGSG          90      ATGGCAGCATGCTTTACGGCTAAGTGCCTTTAGCTTCCGCCTTGTGGCGGTTCTGGCG
MAACNGHQCLSLPPCGGSG          56      ATGGCAGCATGCAATGGTCATCAGTGCTTGTCGTTGCCTCCGTGTGGCGGTTCTGGCG
MAACRPKQCWQLPPCGGSG          1798    ATGGCAGCATGCCGTCCGAAGCAGTGCTGGCAGTTGCCTCCGTGTGGCGGTTCTGGCG
MAACYREQCPHLPPCGGSG          43      ATGGCAGCATGCTATCGTGAGCAGTGCCCTCATCTTCCTCCTTGTGGCGGTTCTGGCG
MAACLYPRCPSLPPCGGSG          465     ATGGCAGCATGCCTGTATCCGCGTTGCCCTTCTCTTCCTCCGTGTGGCGGTTCTGGCG
MAACFKSCCHQLPPCGGSG          39      ATGGCAGCATGCTTTAAGAGTTGTTGCCATTAGCTTCCGCCTTGTGGCGGTTCTGGCG
MAACFWDLCHLLPPCGGSG          92      ATGGCAGCATGCTTTTGGGATCTTTGCCATCTGCTTCCTCCGTGTGGCGGTTCTGGCG

 Group17
MAACLWLGGCHPQSCGGSG          49      ATGGCAGCATGCCTTTGGTTGGGGGGTTGCCATCCGTAGTCTTGTGGCGGTTCTGGCG
MAACHWSAWCHPQNCGGSG          49      ATGGCAGCATGCCATTGGTCGGCGTGGTGCCATCCTCAGAATTGTGGCGGTTCTGGCG
MAACSFIQDCHPQSCGGSG          1574     ATGGCAGCATGCTCTTTTATTTAGGATTGCCATCCTCAGTCGTGTGGCGGTTCTGGCG
MAACTYFSDCHPQHCGGSG          356      ATGGCAGCATGCACGTATTTTTCGGATTGCCATCCTTAGCATTGTGGCGGTTCTGGCG

 Group MIXED
```

MAACFWLLPPCSHSCGGSG	33	ATGGCAGCATGCTTTTGGTTGTTGCCTCCTTGCTCTCATTCGTGTGGCGGTTCTGGCG
MAACMFVLPPCQLECGGSG	33	ATGGCAGCATGCATGTTTGTTTTGCCTCCGTGCTAGCTGGAGTGTGGCGGTTCTGGCG
MAACSVFRCSSESQCGGSG	34	ATGGCAGCATGCTCGGTGTTTAGGTGCTCTTCGGAGTCTTAGTGTGGCGGTTCTGGCG
MAACSILPCPHPVQCGGSG	36	ATGGCAGCATGCTCTATTCTGCCTTGCCCTCATCCTGTTTAGTGTGGCGGTTCTGGCG
MAACPHPQFCYFYNCGGSG	64	ATGGCAGCATGCCCTCATCCTTAGTTTTGCTATTTTTATAATTGTGGCGGTTCTGGCG
MAACPFENVLCYPQCGGSG	39	ATGGCAGCATGCCCGTTTGAGAATGTGTTGTGCTATCCTCAGTGTGGCGGTTCTGGCG
MAACHPQVCPSKSACGGSG	50	ATGGCAGCATGCCATCCTTAGGTGTGCCCGTCTAAGTCGGCGTGTGGCGGTTCTGGCG
MAACPCSASCSVPHCGGSG	62	ATGGCAGCATGCCCGTGTAGTGCTAGTTGCTCTGTTCCTCATTGTGGCGGTTCTGGCG
MAACVPKHCVHPQVCGGSG	35	ATGGCAGCATGCGTGCCTAAGCATTGCGTTCATCCTTAGGTTTGTGGCGGTTCTGGCG
MAACLPPHSCWNQVCGGSG	1227	ATGGCAGCATGCCTGCCTCCGCATTCTTGCTGGAATCAGGTTTGTGGCGGTTCTGGCG
MAACVSRCDNTLFPCGGSG	45	ATGGCAGCATGCGTTTCGCGTTGCGATAATACTCTGTTTCCTTGTGGCGGTTCTGGCG
MAACLSGCQDVLPPCGGSG	81	ATGGCAGCATGCTTGTCGGGTTGCTAGGATGTTTTGCCTCCTTGTGGCGGTTCTGGCG
MAACQGEPACIQYPCGGSG	37	ATGGCAGCATGCTAGGGTGAGCCTGCTTGCATTCAGTATCCGTGTGGCGGTTCTGGCG
MAACSIAYQCLAQHCGGSG	49	ATGGCAGCATGCTCGATAGCTTATCAGTGCTTGGCGTAGCATTGTGGCGGTTCTGGCG
MAACIYYCDCSLNLCGGSG	62	ATGGCAGCATGCATTTATTATTGTGATTGCTCGCTTAATTTGTGTGGCGGTTCTGGCG
MAACTRYQICTASLCGGSG	35	ATGGCAGCATGCACGCGGTATTAGATTTGCACTGCTAGTTTGTGTGGCGGTTCTGGCG
MAACYAVVHCPASSCGGSG	77	ATGGCAGCATGCTATGCGGTGGTTCATTGCCCGGCTAGTTCTTGTGGCGGTTCTGGCG
MAACLKVFNCCLANCGGSG	35	ATGGCAGCATGCCTTAAGGTGTTTAATTGCTGTCTTGCTAATTGTGGCGGTTCTGGCG
MAACSSPRCLYLTACGGSG	36	ATGGCAGCATGCTCGTCGCCTCGTTGCCTGTATCTTACGGCGTGTGGCGGTTCTGGCG
MAACYLPPCDPFILCGGSG	95	ATGGCAGCATGCTATCTTCCTCCGTGCGATCCGTTTATTTTGTGTGGCGGTTCTGGCG
MAACFKKICPPCWKCGGSG	38	ATGGCAGCATGCTTTAAGAAGATTTGCCCGCCTTGTTGGAAGTGTGGCGGTTCTGGCG
MAACPQGSCQAHPNCGGSG	34	ATGGCAGCATGCCCGTAGGGGTCGTGCTAGGCTCATCCTAATTGTGGCGGTTCTGGCG
MAACWRFSCGSQTICGGSG	42	ATGGCAGCATGCTGGAGGTTTTCGTGCGGTTCGTAGACGATTTGTGGCGGTTCTGGCG
MAACVWFLCSSTPVCGGSG	40	ATGGCAGCATGCGTTTGGTTTTTGTGCTCTTCGACTCCGGTTTGTGGCGGTTCTGGCG
MAACVYVVCWQPLNCGGSG	40	ATGGCAGCATGCGTTTATGTTGTTTGCTGGTAGCCTTTGAATTGTGGCGGTTCTGGCG
MAACYITCHPQFFTCGGSG	115	ATGGCAGCATGCTATATTACTTGCCATCCGTAGTTTTTTACTTGTGGCGGTTCTGGCG
MAACPTIPPHCYQACGGSG	50	ATGGCAGCATGCCCGACTATTCCTCCTCATTGCTATTAGGCGTGTGGCGGTTCTGGCG
MAACLPPWGQCFYSCGGSG	47	ATGGCAGCATGCTTGCCGCCTTGGGGTTAGTGCTTTTATAGTTGTGGCGGTTCTGGCG
MAACLYRVSECNTLCGGSG	47	ATGGCAGCATGCCTGTATCGTGTTTCGGAGTGCAATACGTTGTGTGGCGGTTCTGGCG

uPA – Library B

Group1		
MAACAHSWCTARIHCGGSG	127	ATGGCAGCATGCGCTCATTCGTGGTGCACTGCTCGTATTCATTGTGGCGGTTCTGGCG
MAACAHQLCTARAYCGGSG	87	ATGGCAGCATGCGCTCATCAGCTGTGCACTGCGCGTGCTTATTGTGGCGGTTCTGGCG
MAACLHPMCTARSSCGGSG	146	ATGGCAGCATGCTTGCATCCTATGTGCACTGCGCGTTCGTCGTGTGGCGGTTCTGGCG
MAACLQPGCTARVSCGGSG	115	ATGGCAGCATGCCTTCAGCCTGGGTGCACGGCGCGTGTGTCTTGTGGCGGTTCTGGCG
MAACLRPLCSARMHCGGSG	87	ATGGCAGCATGCCTGCGTCCGTTGTGCTCGGCGCGTATGCATTGTGGCGGTTCTGGCG
MAACMGQLCTARSGCGGSG	428	ATGGCAGCATGCATGGGTTAGCTGTGCACGGCGAGGTCTGGTTGTGGCGGTTCTGGCG
MAACMQHRCSARTGCGGSG	166	ATGGCAGCATGCATGTAGCATAGGTGCTCGGCGCGGACTGGTTGTGGCGGTTCTGGCG
MAACSSDNCTARVTCGGSG	109	ATGGCAGCATGCTCGTCTGATAATTGCACGGCTAGGGTTACTTGTGGCGGTTCTGGCG
MAACPAVNCTARTTCGGSG	104	ATGGCAGCATGCCCTGCTGTGAATTGCACTGCGCGTACGACGTGTGGCGGTTCTGGCG
MAACSASQCSARIGCGGSG	1493	ATGGCAGCATGCTCTGCTTCGTAGTGCTCTGCTAGGATTGGTTGTGGCGGTTCTGGCG
MAACSATQCTARLSCGGSG	145	ATGGCAGCATGCTCTGCGACTCAGTGCACTGCGCGTTTGTCGTGTGGCGGTTCTGGCG
MAACKETQCTARITCGGSG	268	ATGGCAGCATGCAAGGAGACGTAGTGCACGGCGCGGATTACTTGTGGCGGTTCTGGCG
MAACFNTQCTARLSCGGSG	99	ATGGCAGCATGCTTTAATACGTAGTGCACTGCGCGTCTTTCTTGTGGCGGTTCTGGCG
MAACRTAVCTARLLCGGSG	215	ATGGCAGCATGCCGTACTGCTGTGTGCACTGCTCGTTTGTTGTGTGGCGGTTCTGGCG
MAACRSAVCTARVRCGGSG	191	ATGGCAGCATGCCGGTCTGCTGTGTGCACTGCGCGGGTGCGTTGTGGCGGTTCTGGCG
MAACAASVCTARLFCGGSG	751	ATGGCAGCATGCGCGGCTTCGGTGTGCACTGCTAGGGTTGTTTTGTGGCGGTTCTGGCG
MAACSAAYCTARLQCGGSG	162	ATGGCAGCATGCAGTGCGGCTTATTGCACGGCTAGGCTGTAGTGTGGCGGTTCTGGCG
MAACRQSTCSARTYCGGSG	493	ATGGCAGCATGCCGGTAGTCTACTTGCTCTGCTAGGACGTATTGTGGCGGTTCTGGCG
MAACKQSVCTARTLCGGSG	262	ATGGCAGCATGCAAGTAGAGTGTGTGCACGGCTAGGACGTTGTGTGGCGGTTCTGGCG
MAACTQSACSARVVCGGSG	5410	ATGGCAGCATGCACGTAGTCTGCTTGCTCGGCGAGGGTTGTGTGTGGCGGTTCTGGCG
MAACLESSCSARIVCGGSG	108	ATGGCAGCATGCCTTGAGAGTTCTTGCTCGGCTCGTATTGTTTGTGGCGGTTCTGGCG
MAACNESVCSARKQCGGSG	148	ATGGCAGCATGCAATGAGTCGGTGTGCTCTGCGCGTAAGTAGTGTGGCGGTTCTGGCG
MAACYGSACSARSSCGGSG	266	ATGGCAGCATGCTATGGGTCTGCGTGCAGTGCGCGTTCTTCTTGTGGCGGTTCTGGCG
MAACFNSACTARSMCGGSG	238	ATGGCAGCATGCTTTAATTCGGCTTGCACTGCTCGGTCGATGTGTGGCGGTTCTGGCG
MAACTYALCTARTFCGGSG	1905	ATGGCAGCATGCACTTATGCTCTGTGCACTGCGCGTACGTTTTGTGGCGGTTCTGGCG
MAACNFSLCSARFFCGGSG	96	ATGGCAGCATGCAATTTTTCTCTTTGCTCGGCGCGTTTTTTTTTGTGGCGGTTCTGGCG
MAACSVSFCSARSFCGGSG	103	ATGGCAGCATGCAGTGTGTCTTTTTGCTCTGCGCGGTCTTTTTGTGGCGGTTCTGGCG
MAACSLASCSARMLCGGSG	413	ATGGCAGCATGCTCTCTGGCGAGTTGCTCGGCGCGTATGTTGTGTGGCGGTTCTGGCG
MAACTLGNCTARAICGGSG	195	ATGGCAGCATGCACTCTTGGTAATTGCACGGCGAGGGCTATTTGTGGCGGTTCTGGCG
MAACPLSACSGRTLCGGSG	1026	ATGGCAGCATGCCCGCTTTCTGCGTGCTCGGGGAGGACGTTGTGTGGCGGTTCTGGCG
MAACPVRSCTARQACGGSG	298	ATGGCAGCATGCCCTGTGCGTAGTTGCACGGCTCGTTAGGCGTGTGGCGGTTCTGGCG
MAACPETSCTARQVCGGSG	120	ATGGCAGCATGCCCTGAGACGTCTTGCACTGCGCGTTAGGTTTGTGGCGGTTCTGGCG
MAACPQASCSARRYCGGSG	165	ATGGCAGCATGCCCTCAGGCGTCTTGCTCTGCGCGTCGGTATTGTGGCGGTTCTGGCG
MAACLPSLCSARSRCGGSG	179	ATGGCAGCATGCCTGCCGAGTCTTTGCTCGGCGAGGTCTCGTTGTGGCGGTTCTGGCG
MAACVPTQCSARSSCGGSG	117	ATGGCAGCATGCGTTCCGACGCAGTGCTCGGCGCGTTCGAGTTGTGGCGGTTCTGGCG
MAACSPSQCTARAGCGGSG	799	ATGGCAGCATGCTCTCCTTCGTAGTGCACTGCTCGTGCGGGGTGTGGCGGTTCTGGCG
MAACSPSMCTARVACGGSG	173	ATGGCAGCATGCAGTCCTTCGATGTGCACTGCGCGTGTTGCGTGTGGCGGTTCTGGCG
MAACSPSMCSGRRSCGGSG	449	ATGGCAGCATGCTCGCCGTCGATGTGCAGTGGGCGTAGGTCTTGTGGCGGTTCTGGCG
MAACSPSTCSGRSLCGGSG	107	ATGGCAGCATGCTCGCCGTCGACGTGCTCTGGGCGTTCTTTGTGTGGCGGTTCTGGCG
MAACRPALCSGRTACGGSG	99	ATGGCAGCATGCCGTCCTGCGCTTTGCAGTGGTCGGACTGCTTGTGGCGGTTCTGGCG
MAACAPSECSARFLCGGSG	192	ATGGCAGCATGCGCCGCCGTCAGAGTGCTCTGCTCGGTTTTTGTGTGGCGGTTCTGGCG
MAACAPFTCSARFTCGGSG	111	ATGGCAGCATGCGCTCCTTTTACGTGCTCTGCGCGTTTTACGTGTGGCGGTTCTGGCG

MAACVPHSCTARTFCGGSG	330	ATGGCAGCATGCGTGCCTCATTCGTGCACTGCTCGTACTTTTTGTGGCGGTTCTGGCG
MAACLPTSCTARLWCGGSG	261	ATGGCAGCATGCCTGCCTACGTCTTGCACTGCGCGTTTGTGGTGTGGCGGTTCTGGCG
MAACMPSACTRRDFCGGSG	339	ATGGCAGCATGCATGCCTTCTGCGTGCACGCGGCGTGATTTTTGTGGCGGTTCTGGCG
MAACQPSRCSARDHCGGSG	248	ATGGCAGCATGCCAGCCGAGTAGGTGCAGTGCTCGGGATCATTGTGGCGGTTCTGGCG
MAACWTPTCSARSHCGGSG	890	ATGGCAGCATGCTGGACGCCTACGTGCTCTGCTCGTTCGCATTGTGGCGGTTCTGGCG
MAACWMASCSARSDCGGSG	221	ATGGCAGCATGCTGGATGGCTTCGTGCTCTGCGCGTTCTGATTGTGGCGGTTCTGGCG
MAACGVVTCTARQHCGGSG	683	ATGGCAGCATGCGGTGTTGTGACGTGCACGGCTCGTTAGCATTGTGGCGGTTCTGGCG
MAACGIANCTARAQCGGSG	162	ATGGCAGCATGCGGGATTGCTAATTGCACGGCGCGTGCTTAGTGTGGCGGTTCTGGCG
Group2		
MAACAQATSCQTARCGGSG	805	ATGGCAGCATGCGCGTAGGCGACTTCGTGCTAGACTGCGCGTTGTGGCGGTTCTGGCG
MAACSVVTSCLTARCGGSG	147	ATGGCAGCATGCTCTGTTGTGACTTCTTGCTTGACGGCTCGGTGTGGCGGTTCTGGCG
MAACITYSSCSTARCGGSG	236	ATGGCAGCATGCATTACGTATTCGTCGTGCTCGACTGCTCGGTGTGGCGGTTCTGGCG
MAACSSYDACISARCGGSG	437	ATGGCAGCATGCTCGTCTTATGATGCTTGCATTTCTGCGAGGTGTGGCGGTTCTGGCG
MAACLSLRSCLSARCGGSG	159	ATGGCAGCATGCCTTTCTTTGCGGTCGTGCCTGTCTGCTAGGTGTGGCGGTTCTGGCG
MAACRPMRSCFSARCGGSG	111	ATGGCAGCATGCAGGCCTATGCGGTCGTGCTTTTCTGCGAGGTGTGGCGGTTCTGGCG
MAACFEQESCLSARCGGSG	165	ATGGCAGCATGCTTTGAGCAGGAGTCTTGCCTTTCTGCTCGGTGTGGCGGTTCTGGCG
MAACRLDLSCLSARCGGSG	408	ATGGCAGCATGCCGTTTGGATTTGAGTTGCTTGTCGGCGCGTTGTGGCGGTTCTGGCG
MAACQVNLACTSARCGGSG	339	ATGGCAGCATGCTAGGTGAATCTTGCGTGCACTTCGGCTAGGTGTGGCGGTTCTGGCG
MAACSYTLSCLSGRCGGSG	181	ATGGCAGCATGCTCGTATACTTTGAGTTGCTTGTCTGGTCGGTGTGGCGGTTCTGGCG
MAACMLSCLTGRCGGSG	135	ATGGCAGCATGCATGCTGTCGTGCTTGACGGGGCGGTGTGGCGGTTCTGGCG
MAACILSLPCVSARCGGSG	168	ATGGCAGCATGCATTTTGTCGTTGCCTTGCGTTAGTGCGCGTTGTGGCGGTTCTGGCG
MAACLLMAPCQTARCGGSG	103	ATGGCAGCATGCTTGCTTATGGCGCCGTGCTAGACTGCGAGGTGTGGCGGTTCTGGCG
MAACLWRMPCLTARCGGSG	238	ATGGCAGCATGCCTGTGGAGGATGCCTTGCCTGACTGCTCGTTGTGGCGGTTCTGGCG
MAACALRVPCFTGRCGGSG	380	ATGGCAGCATGCGCTCTGAGGGTTCCTTGCTTTACTGGGCGGTGTGGCGGTTCTGGCG
MAACVTVYPCTTARCGGSG	417	ATGGCAGCATGCGTTACTGTGTATCCGTGCACGACGGCGAGGTGTGGCGGTTCTGGCG
MAACRALYPCGTARCGGSG	193	ATGGCAGCATGCCGTGCGTTGTATCCTTGCGGGACGGCTCGTTGTGGCGGTTCTGGCG
MAACTARQPCSTARCGGSG	185	ATGGCAGCATGCACTGCTAGGCAGCCGTGCTCGACGGCGCGTTGTGGCGGTTCTGGCG
Group3		
MAACATARCMQAYLCGGSG	347	ATGGCAGCATGCGCGACTGCTAGGTGCATGTAGGCTTATTTGTGTGGCGGTTCTGGCG
MAACLTARCQSSFLCGGSG	109	ATGGCAGCATGCCTGACTGCTAGGTGCTAGTCTTCGTTTTTGTGTGGCGGTTCTGGCG
MAACPTARCPQSVLCGGSG	118	ATGGCAGCATGCCCTACTGCTCGTTGCCCTTAGTCTGTTTTTGTGTGGCGGTTCTGGCG
MAACSTARCLLSYQCGGSG	264	ATGGCAGCATGCTCGACGGCTCGGTGCCTGTTGTCTTATCAGTGTGGCGGTTCTGGCG
MAACSTARCPLSYACGGSG	94	ATGGCAGCATGCTCTACGGCGCGTTGCCCGTTGTCTTATGCGTGTGGCGGTTCTGGCG
MAACSSTARCELSYCGGSG	116	ATGGCAGCATGCTCGTCTACTGCGCGTTGCGAGCTTTCGTATTGTGGCGGTTCTGGCG
MAACATARCSLPSLCGGSG	200	ATGGCAGCATGCGCGACTGCTCGTTGCTCTCTGCCGTCTCTTTGTGGCGGTTCTGGCG
MAACETARCSFYSLCGGSG	124	ATGGCAGCATGCGAGACTGCTAGGTGCTCGTTTTATTCGCTGTGTGGCGGTTCTGGCG
MAACSTARCSDRSMCGGSG	404	ATGGCAGCATGCTCGACTGCGCGTTGCTCGGATCGTTCGATGTGTGGCGGTTCTGGCG
MAACPTARCRVPQLCGGSG	95	ATGGCAGCATGCCCGACGGCTCGGTGCCGGGTTCCTTAGTTGTGTGGCGGTTCTGGCG
MAACNTARCFPSSWCGGSG	200	ATGGCAGCATGCAATACTGCGCGTTGCTTTCCGTCGAGTTGGTGTGGCGGTTCTGGCG
MAACSTARCTPTEFCGGSG	122	ATGGCAGCATGCTCGACTGCGAGGTGCACTCCTACGGAGTTTTGTGGCGGTTCTGGCG
MAACPTARCIPWTSCGGSG	188	ATGGCAGCATGCCCTACGGCTAGGTGCATTCCTTGGACTTCTTGTGGCGGTTCTGGCG
MAACATARCLWTMTCGGSG	138	ATGGCAGCATGCGCGACGGCGCGTTGCCTGTGGACTATGACTTGTGGCGGTTCTGGCG
Group4		
MAACSYSCLSARTSCGGSG	667	ATGGCAGCATGCTCGTATTCGTGCTTGTCTGCTAGGACTAGTTGTGGCGGTTCTGGCG
MAACTLACLSARGSCGGSG	297	ATGGCAGCATGCACTTTGGCGTGCCTGAGTGCTAGGGGTTCGTGTGGCGGTTCTGGCG
MAACQFSCISARSFCGGSG	399	ATGGCAGCATGCTAGTTTTCGTGCATTTCGGCGCGGTCTTTTTGTGGCGGTTCTGGCG
MAACVYACFTARSQCGGSG	289	ATGGCAGCATGCGTGTATGCGTGCTTTACTGCTCGGTCTTAGTGTGGCGGTTCTGGCG
MAACNLSCYTGRSLCGGSG	182	ATGGCAGCATGCAATCTATCGTGCTATACTGGGCGTTCTCTTTGTGGCGGTTCTGGCG
MAACLVGCTTARTFCGGSG	163	ATGGCAGCATGCCTGGTGGGGTGCACTACTGCTCGGACGTTTTGTGGCGGTTCTGGCG
Group5		
MAACPVLPQCSARSCGGSG	541	ATGGCAGCATGCCCTGTGCTTCCTTAGTGCTCGGCGAGGTCTTGTGGCGGTTCTGGCG
MAACPSSPRCTARSCGGSG	249	ATGGCAGCATGCCCTTCTTCGCCGCGTTGCACTGCTAGGAGTTGTGGCGGTTCTGGCG
MAACIVNPICTARTCGGSG	138	ATGGCAGCATGCATTGTGAATCCTATTTGCACTGCGCGTACGTGTGGCGGTTCTGGCG
MAACMLSGSCTARSCGGSG	1381	ATGGCAGCATGCATGTTGTCTGGTTCTTGCACGGCTAGGTCGTGTGGCGGTTCTGGCG
MAACTLKNNCTARYCGGSG	130	ATGGCAGCATGCACGCTGAAGAATAATTGCACTGCTCGTTATTGTGGCGGTTCTGGCG
MAACLPQSQCTARYCGGSG	219	ATGGCAGCATGCCTTCCGCAGTCTCAGTGCACTGCTCGGTATTGTGGCGGTTCTGGCG
MAACGPQPDCTARYCGGSG	157	ATGGCAGCATGCGGGCCGTAGCCTGATTGCACTGCTCGTTATTGTGGCGGTTCTGGCG
Group6		
MAACLQFCSSQSARCGGSG	117	ATGGCAGCATGCCTGTAGTTTTGCTCGTCTCAGTCGGCTCGTTGTGGCGGTTCTGGCG
MAACLWYCAAESARCGGSG	106	ATGGCAGCATGCCTTTGGTATTGCGCGGCTGAGTCTGCGCGTTGTGGCGGTTCTGGCG
MAACFQTCTLQTARCGGSG	365	ATGGCAGCATGCTTTCAGACGTGCACGCTTTAGACTGCTCGTTGTGGCGGTTCTGGCG
MAACRSVCAVATARCGGSG	109	ATGGCAGCATGCCGTTCGGTGTGCGCTGTGGCGACGGCGAGGTGTGGCGGTTCTGGCG
MAACGLLCSSLSARCGGSG	340	ATGGCAGCATGCGGGTTGCTTTGCTCTTCTTTGTCGGCGCGTTGTGGCGGTTCTGGCG
Group7		
MAACAPDQCTKFTMCGGSG	2493	ATGGCAGCATGCGCTCCGGATCAGTGCACTAAGTTTACTATGTGTGGCGGTTCTGGCG
MAACAQTSCSKYSLCGGSG	112	ATGGCAGCATGCGCGTAGACGTCTTGCTCGAAGTATTCGCTGTGTGGCGGTTCTGGCG
MAACPAQNCTKYTLCGGSG	447	ATGGCAGCATGCCCTGCTCAGAATTGCACTAAGTATACTTTGTGTGGCGGTTCTGGCG
MAACPTRLCTKFTLCGGSG	93	ATGGCAGCATGCCCGACGCGGTTGTGCACGAAGTTTACGTTGTGTGGCGGTTCTGGCG
MAACPPVMCQRFALCGGSG	442	ATGGCAGCATGCCCTCCTGTTATGTGCCAGCGGTTTGCGTTGTGTGGCGGTTCTGGCG
MAACSPVYCEKFTLCGGSG	170	ATGGCAGCATGCTCGCCGGTGTATTGCGAGAAGTTTACTTTGTGTGGCGGTTCTGGCG
Group8		

MAACFRYQCTARSHCGGSG	1081	ATGGCAGCATGCTTTCGTTATCAGTGCACGGCGCGTTCTCATTGTGGCGGTTCTGGCG
MAACFTYLCSARHHCGGSG	90	ATGGCAGCATGCTTTACTTATCTTTGCTCGGCTCGGCATCATTGTGGCGGTTCTGGCG
MAACKDYVCSARLHCGGSG	105	ATGGCAGCATGCAAGGATTATGTGTGCTCGGCTCGTTTGCATTGTGGCGGTTCTGGCG
MAACFDFLCSARTCGGSG	128	ATGGCAGCATGCTTTGATTTCCTGTGCTCGGCTAGGACTTGTGGCGGTTCTGGCG
Group9		
MAACKHSDCTARFPCGGSG	1315	ATGGCAGCATGCAAGCATAGTGATTGCACTGCTCGGTTTCCTTGTGGCGGTTCTGGCG
MAACTSSDCTRRYPCGGSG	121	ATGGCAGCATGCACGAGTTCGGATTGCACGCGGCGTTATCCTTGTGGCGGTTCTGGCG
MAACVLSDCSARLPCGGSG	87	ATGGCAGCATGCGTGCTGTCGGATTGCTCTGCGAGGTTGCCTTGTGGCGGTTCTGGCG
MAACPLQLCTARYPCGGSG	761	ATGGCAGCATGCCCTTTGTAGTTGTGCACGGCTCGGTATCCTTGTGGCGGTTCTGGCG
MAACSLSLCSARYPCGGSG	166	ATGGCAGCATGCTCTCTGAGTCTTTGCTCTGCGCGTTATCCGTGTGGCGGTTCTGGCG
Group10		
MAACLKRCSGTARCGGSG	256	ATGGCAGCATGCCTGAAGCGTTGCTCTGGTACTGCTCGTTGTGGCGGTTCTGGCG
MAACAYRSCQGTARCGGSG	918	ATGGCAGCATGCGCTTATCGGTCGTGCTAGGGGACGGCTAGGTGTGGCGGTTCTGGCG
MAACYWSSCTGTARCGGSG	189	ATGGCAGCATGCTATTGGTCGTCTTGCACTGGGACGGCGAGGTGTGGCGGTTCTGGCG
MAACMWQVCASTARCGGSG	97	ATGGCAGCATGCATGTGGCAGGTGTGCGCTTCGACGGCGCGGTGTGGCGGTTCTGGCG
MAACYAICSSTARSCGGSG	326	ATGGCAGCATGCTATGCTATTTGCAGTTCGACGGCTCGGTCTTGTGGCGGTTCTGGCG
MAACYMHFCHTTARCGGSG	101	ATGGCAGCATGCTATATGCATTTTTGCCATACGACGGCTCGTTGTGGCGGTTCTGGCG
MAACNPYLCNPTARCGGSG	93	ATGGCAGCATGCAATCCGTATCTGTGCAATCCGACTGCTAGGTGTGGCGGTTCTGGCG
MAACSPLRCTLTARCGGSG	156	ATGGCAGCATGCAGTCCTCTTCGTTGCACTTTGACGGCGCGGTGTGGCGGTTCTGGCG
MAACTARLCQQSARCGGSG	95	ATGGCAGCATGCACGGCGAGGCTTTGCTAGCAGTCGGCTCGTTGTGGCGGTTCTGGCG
Group11		
MAACSLSTARCPMQCGGSG	348	ATGGCAGCATGCTCTTTGAGTACGGCGCGGTGCCCTATGCAGTGTGGCGGTTCTGGCG
MAACLSTARCPMQPCGGSG	164	ATGGCAGCATGCCTGTCGACGGCGCGTTGCCCGATGTAGCCGTGTGGCGGTTCTGGCG
MAACSSSARHCPLVCGGSG	150	ATGGCAGCATGCTCTTCGAGTGCTCGTCATTGCCCTCTTGTGTGTGGCGGTTCTGGCG
MAACVVTARCPAFHCGGSG	181	ATGGCAGCATGCGTGGTTACGGCTAGGTGCCCGGCGTTTCATTGTGGCGGTTCTGGCG
MAACNSSARCPSFTCGGSG	135	ATGGCAGCATGCAATTCTTCTGCGCGGTGCCCTTCGTTTACGTGTGGCGGTTCTGGCG
MAACVTARCPSYPFCGGSG	104	ATGGCAGCATGCGTGACGGCTCGTTGCCCGAGTTATCCGTTTTGTGGCGGTTCTGGCG
Group12		
MAACNVPLCTARLSCGGSG	778	ATGGCAGCATGCAATGTTCCGTTGTGCACTGCTCGTTTGTCGTGTGGCGGTTCTGGCG
MAACLAGWCSARQACGGSG	120	ATGGCAGCATGCTTGGCTGGTTGGTGCAGTGCGAGGTAGGCGTGTGGCGGTTCTGGCG
MAACLTGLCTARGYCGGSG	728	ATGGCAGCATGCTTGACGGGTCTGTGCACTGCTCGTGGTTATTGTGGCGGTTCTGGCG
MAACVTSGCTARWWCGGSG	88	ATGGCAGCATGCGTTACTAGTGGGTGCACGGCGCGGTGGTGGTGTGGCGGTTCTGGCG
MAACKSSLCSARQWCGGSG	251	ATGGCAGCATGCAAGTCTTCTCTGTGCAGTGCTCGGCAGTGGTGTGGCGGTTCTGGCG
MAACSYAHCSARWTCGGSG	334	ATGGCAGCATGCTCTTATGCTCATTGCAGTGCTAGGTGGACGTGTGGCGGTTCTGGCG
MAACLESHCSARWSCGGSG	264	ATGGCAGCATGCTTGGAGTCGCATTGCTCGGCGCGTTGGTCGTGTGGCGGTTCTGGCG
MAACRHLSCTARTNCGGSG	269	ATGGCAGCATGCAGGCATCTGTCTTGCACGGCTCGTACTAATTGTGGCGGTTCTGGCG
Group13		
MAACDNRAGCSTARCGGSG	195	ATGGCAGCATGCGATAATAGGGCTGGTTGCAGTACTGCTCGGTGTGGCGGTTCTGGCG
MAACMNRPSCNSARCGGSG	355	ATGGCAGCATGCATGAATAGGCCTAGTTGCAATTCTGCGCGTTGTGGCGGTTCTGGCG
MAACVPKLACESARCGGSG	255	ATGGCAGCATGCGTTCCTAAGTTGGCTTGCGAGTCTGCGCGTTGTGGCGGTTCTGGCG
MAACSLMPFCYSARCGGSG	164	ATGGCAGCATGCAGTTTGATGCCTTTTTGCTATTCTGCGCGTTGTGGCGGTTCTGGCG
MAACKQQPMCATARCGGSG	460	ATGGCAGCATGCAAGCAGTAGCCGATGTGCGCGACGGCTAGGTGTGGCGGTTCTGGCG
MAACPPPPSCFTARCGGSG	392	ATGGCAGCATGCCCTCCTCCGCCGTCGTGCTTTACGGCTAGGTGTGGCGGTTCTGGCG
MAACPVKPSCHSGRCGGSG	1056	ATGGCAGCATGCCCTGTTAAGCCGTCTTGCCATTCTGGGAGGTGTGGCGGTTCTGGCG
MAACFHLQHCATARCGGSG	98	ATGGCAGCATGCTTTCATCTTTAGCATTGCGCTACGGCTCGTTGTGGCGGTTCTGGCG
Group14		
MAACNALFSGCAYMCGGSG	263	ATGGCAGCATGCAATGCTCTTTTTTCGGGTTGCGCTTATATGTGTGGCGGTTCTGGCG
MAACTPYYSLCTHPCGGSG	254	ATGGCAGCATGCACGCCTTATTATTCGCTGTGCACGCATCCGTGTGGCGGTTCTGGCG
MAACKPLFSLCRAPCGGSG	130	ATGGCAGCATGCAAGCCGCTGTTTTCGTTGTGCCGTGCGCCGTGTGGCGGTTCTGGCG
Group15		
MAACSLNSCFSARSCGGSG	336	ATGGCAGCATGCTCGTTGAATTCTTGCTTTTCGGCTCGTTCTTGTGGCGGTTCTGGCG
MAACQYPSCMTARSCGGSG	108	ATGGCAGCATGCCAGTATCCGTCTTGCATGACTGCTAGGTCTTGTGGCGGTTCTGGCG
MAACLYSSCPTARYCGGSG	162	ATGGCAGCATGCCTGTATTCGTCGTGCCCGACTGCGCGTTATTGTGGCGGTTCTGGCG
MAACLQISCSTARMCGGSG	87	ATGGCAGCATGCCTTTAGATTAGTTGCTCTACGGCTCGGATGTGTGGCGGTTCTGGCG
MAACRQACATARANCGGSG	556	ATGGCAGCATGCAGGCAGGCTTGCGCGACTGCGCGGGCTAATTGTGGCGGTTCTGGCG
MAACTQPCLTARATCGGSG	274	ATGGCAGCATGCACGTAGCCTTGCCTGACTGCTCGGGCTACTTGTGGCGGTTCTGGCG
MAACRRDPCATARNCGGSG	544	ATGGCAGCATGCCGTCGGGATCCGTGCGCGACTGCGAGGAATTGTGGCGGTTCTGGCG
Group16		
MAACSTARCWTHSPCGGSG	111	ATGGCAGCATGCTCGACTGCGCGTTGCTGGACGCATTCTCCTTGTGGCGGTTCTGGCG
MAACPTARCTQLPPCGGSG	143	ATGGCAGCATGCCCTACGGCGCGGTGCACGTAGTTGCCTCCTTGTGGCGGTTCTGGCG
MAACLTARCQFATACGGSG	165	ATGGCAGCATGCTTGACTGCTAGGTGCTAGTTTGCTACTGCTTGTGGCGGTTCTGGCG
MAACSTARCRYLANCGGSG	223	ATGGCAGCATGCTCTACTGCGCGGTGCCGTTATCTGGCGAATTGTGGCGGTTCTGGCG
MAACLTARCFDTPNCGGSG	751	ATGGCAGCATGCTTGACTGCGAGGTGCTTTGATACTCCGAATTGTGGCGGTTCTGGCG
MAACGTARCHALFSCGGSG	396	ATGGCAGCATGCGGGACTGCGAGGTGCCATGCTCTTTTTTCTTGTGGCGGTTCTGGCG
MAACGTGRCSTYFSCGGSG	263	ATGGCAGCATGCGGGACTGGGCGGTGCTCTACGTATTTTTCTTGTGGCGGTTCTGGCG
MAACETARCLALWACGGSG	126	ATGGCAGCATGCGAGACGGCTCGGTGCCTGGCTTTGTGGGCTTGTGGCGGTTCTGGCG
Group17		
MAACHKLMNCRFSLCGGSG	781	ATGGCAGCATGCCATAAGTTGATGAATTGCCGTTTTTCGCTTTGTGGCGGTTCTGGCG
MAACSYMGNCKFSLCGGSG	193	ATGGCAGCATGCTCGTATATGGGTAATTGCAAGTTTTCGTTGTGTGGCGGTTCTGGCG

```
MAACIFYKSCKYSLCGGSG          588      ATGGCAGCATGCATTTTTTATAAGTCTTGCAAGTATTCGTTGTGTGGCGGTTCTGGCG

 Group MIXED
MAACAPPWYLTYSICGGSG          147      ATGGCAGCATGCGCGCCGCCTTGGTACTTGACGTATTCGATTTGTGGCGGTTCTGGCG
MAACCQWWDCSRGLCGGSG          529      ATGGCAGCATGCTGTCAGTGGTGGGATTGCTCTCGTGGTTTGTGTGGCGGTTCTGGCG
MAACMDSCLSRRSHCGGSG          111      ATGGCAGCATGCATGGATTCTTGCTTGTCTCGTCGTTCGCATTGTGGCGGTTCTGGCG
MAACATSCTARMQYCGGSG          309      ATGGCAGCATGCGCTACTTCGTGCACTGCTAGGATGTAGTATTGTGGCGGTTCTGGCG
MAACTARYHCCSFGCGGSG          140      ATGGCAGCATGCACGGCGCGGTATCATTGCTGTAGTTTTGGTTGTGGCGGTTCTGGCG
MAACTARAHCPSQMCGGSG          101      ATGGCAGCATGCACGGCGCGTGCTCATTGCCCGTCTTAGATGTGTGGCGGTTCTGGCG
MAACGDCLSCSTARCGGSG          156      ATGGCAGCATGCGGGGATTGTCTGTCTTGCTCGACGGCGCGGTGTGGCGGTTCTGGCG
MAACGSFFPCFSARCGGSG          95       ATGGCAGCATGCGGGAGTTTTTTTCCTTGCTTTAGTGCTCGGTGTGGCGGTTCTGGCG
MAACPVSCWNETARCGGSG          118      ATGGCAGCATGCCCTGTGTCTTGCTGGAATGAGACGGCGAGGTGTGGCGGTTCTGGCG
MAACPVSCEVTARTCGGSG          109      ATGGCAGCATGCCCGGTGTCTTGCGAGGTTACGGCTCGTACGTGTGGCGGTTCTGGCG
MAACPQVTARCSTMCGGSG          179      ATGGCAGCATGCCCGTAGGTGACTGCGAGGTGCTCTACGATGTGTGGCGGTTCTGGCG
MAACPSARTRCESICGGSG          149      ATGGCAGCATGCCCTTCTGCGCGGACGCGTTGCGAGTCGATTTGTGGCGGTTCTGGCG
MAACTSWFDCTARFCGGSG          117      ATGGCAGCATGCACTTCTTGGTTTGATTGCACGGCTCGTTTTTGTGGCGGTTCTGGCG
MAACSWYTCTARLACGGSG          99       ATGGCAGCATGCTCGTGGTATACTTGCACTGCGCGGTTGGCTTGTGGCGGTTCTGGCG
MAACESMSALCTARCGGSG          97       ATGGCAGCATGCGAGTCTATGTCGGCGTTGTGCACTGCGAGGTGTGGCGGTTCTGGCG
MAACSKIASDCTARCGGSG          128      ATGGCAGCATGCTCGAAGATTGCTTCGGATTGCACGGCTCGGTGTGGCGGTTCTGGCG
MAACPFPPCYMPYPCGGSG          105      ATGGCAGCATGCCCTTTTCCTCCGTGCTATATGCCGATCCTTGTGGCGGTTCTGGCG
MAACFAWDCARYLYCGGSG          93       ATGGCAGCATGCTTTGCGTGGGATTGCGCGCGTTATTTGTATTGTGGCGGTTCTGGCG
MAACNAYYKLCVVQCGGSG          138      ATGGCAGCATGCAATGCGTATTATAAGCTTTGCGTTGTGTAGTGTGGCGGTTCTGGCG
MAACLLSTARCWSSCGGSG          148      ATGGCAGCATGCCTGCTGTCGACTGCGAGGTGCTGGTCTTCGTGTGGCGGTTCTGGCG
MAACTTARSGCWFQCGGSG          104      ATGGCAGCATGCACGACTGCGCGGTCTGGTTGCTGGTTTCAGTGTGGCGGTTCTGGCG
MAACCKRHCPSSLTCGGSG          126      ATGGCAGCATGCTGTAAGCGGCATTGCCCGTCGTCGCTTACTTGTGGCGGTTCTGGCG
MAACSLRSCPFTQFCGGSG          378      ATGGCAGCATGCTCGCTGCGTTCGTGCCCGTTTACGTAGTTTTGTGGCGGTTCTGGCG
MAACEFRCCNPYYSCGGSG          565      ATGGCAGCATGCGAGTTTCGGTGTTGCAATCCTTATTATTCGTGTGGCGGTTCTGGCG
MAACSGRCTHPTGRCGGSG          210      ATGGCAGCATGCTCGGGGAGGTGCACGCATCCTACGGGTAGGTGTGGCGGTTCTGGCG
MAACQQLCTARSTCCGGSG          666      ATGGCAGCATGCTAGCAGCTTTGCACTGCTCGTTCGACGTGTTGTGGCGGTTCTGGCG
MAACQQQQTCTARTCGGSG          455      ATGGCAGCATGCTAGCAGCAGCAGACTTGCACGGCTCGTACTTGTGGCGGTTCTGGCG
MAACRMMECTARGWCGGSG          278      ATGGCAGCATGCCGTATGATGGAGTGCACGGCTAGGGGGTGGTGTGGCGGTTCTGGCG
MAACVISTCSARHDCGGSG          129      ATGGCAGCATGCGTTATTAGTACGTGCAGTGCGCGGCATGATTGTGGCGGTTCTGGCG
MAACVARPSFCSARCGGSG          139      ATGGCAGCATGCGTTGCGCGGCCGTCTTTTTGCTCGGCGCGGTGTGGCGGTTCTGGCG
MAACLPVSFCSARSCGGSG          91       ATGGCAGCATGCCTTCCGGTTTCGTTTTGCTCTGCTCGTTCGTGTGGCGGTTCTGGCG
MAACLPSSHCLTARCGGSG          98       ATGGCAGCATGCCTTCCTTCTTCTCATTGCCTGACTGCTAGGTGTGGCGGTTCTGGCG
MAACKYSLCFYSSSCGGSG          139      ATGGCAGCATGCAAGTATTCTTTGTGCTTTTATAGTAGTTCGTGTGGCGGTTCTGGCG
```

FXIIa – Library 4×4

```
 Group1
MAVCDQWACGQEWCVSHS           381      ATGGCAGTGTGTGATCAGTGGGCGTGCGGTTAGGAGTGGTGTGTGTCTCACTCCG
MAFCSLQKCGQDWCPSHS           90       ATGGCATTTTGTAGTCTGCAGAAGTGCGGTTAGGATTGGTGTCCGTCTCACTCCG
MAWCVQGKCIHAVCASHS           99       ATGGCATGGTGTGTTTAGGGGAAGTGCATTCATGCTGTTTGTGCTTCTCACTCCG
MAWCDLGRCIEGGCLSHS           85       ATGGCATGGTGTGATCTGGGTCGTTGCATTGAGGGTGGTTGTTTGTCTCACTCCG
MAWCANYQCISETCMSHS           86       ATGGCATGGTGTGCTAATTATTAGTGCATTTCGGAGACTTGTATGTCTCACTCCG
MAWCAHLRCRHVGCPSHS           102      ATGGCATGGTGTGCGCATCTGAGGTGCCGGCATGTTGGGTGTCCTTCTCACTCCG
MAWCQGQVCMQCRCFSHS           91       ATGGCATGGTGTTAGGGTCAGGTGTGCATGCAGTGTCGGTGTTTTTCTCACTCCG
MAWCSLSVCQHCTCWSHS           80       ATGGCATGGTGTAGTCTGAGTGTGTGCTAGCATTGTACTTGTTGGTCTCACTCCG
MAWCARDGCRSCACRSHS           94       ATGGCATGGTGTGCTCGTGATGGTTGCCGGAGTTGTGCTTGTCGTTCTCACTCCG

 Group2
MAGCMSLACLQDMCWSHS           85       ATGGCAGGGTGTATGTCGCTGGCGTGCCTGTAGGATATGTGTTGGTCTCACTCCG
MASCGCALCSAQACVSHS           123      ATGGCAAGTTGTGGTTGTGCGTTGTGCTCTGCGTAGGCGTGTGTTTCTCACTCCG
MAQCNCGVCSPQDCLSHS           77       ATGGCACAGTGTAATTGTGGTGTGTGCTCTCCGTAGGATTGTCTGTCTCACTCCG
MACCGCMLCSHLRCQSHS           248      ATGGCATGTTGTGGTTGTATGTTGTGCTCGCATCTGCGGTGTTAGTCTCACTCCG
MALCGCRWCTSWGCPSHS           98       ATGGCATTGTGTGGGTGTAGGTGGTGCACGTCGTGGGGTTGTCCTTCTCACTCCG
MAACCGCRCTSPTCPSHS           94       ATGGCAGCTTGTTGTGGTTGTAGGTGCACGAGTCCGACTTGTCCGTCTCACTCCG

 Group3
MASCTRSWCVTSRCDSHS           280      ATGGCAAGTTGTACTCGTAGTTGGTGCGTTACTTCTAGGTGTGATTCTCACTCCG
MARCVPDWCISKRCRSHS           123      ATGGCAAGGTGTGTTCCTGATTGGTGCATTTCTAAGCGGTGTAGGTCTCACTCCG
MAGCPTAWCLVELCLSHS           175      ATGGCAGGTTGTCCGACGGCTTGGTGCCTTGTGGAGTTGTGTCTGTCTCACTCCG
MANCVSGWCLVRQCFSHS           79       ATGGCAAATTGTGTTTCTGGTTGGTGCCTTGTGCGTCAGTGTTTTTCTCACTCCG
MAQCGLSLCSWNACASHS           79       ATGGCATAGTGTGGTCTGTCGTTGTGCTCTTGGAATGCGTGTGCGTCTCACTCCG
MAQCVVSACSWTKCISHS           78       ATGGCATAGTGTGTGGTGAGTGCGTGCTCTTGGACTAAGTGTATTTCTCACTCCG
MASCYVATCSWWQCGSHS           94       ATGGCATCTTGTTATGTTGCTACTTGCTCTTGGTGGTAGTGTGGGTCTCACTCCG
MASCLLGSCKWQLCQSHS           127      ATGGCATCTTGTCTGCTGGGTAGTTGCAAGTGGTAGCTGTGTCAGTCTCACTCCG
MARCTGRSCEWHACGSHS           369      ATGGCAAGGTGTACGGGGCGGTCGTGCGAGTGGCATGCTTGTGGTTCTCACTCCG

 Group4
MAVCEGTQCSLQCCSSHS           252       ATGGCAGTGTGTGAGGGTACGCAGTGCAGTTTGTAGTGTTGTTCGTCTCACTCCG
MACCEGTSCYVTECQSHS           123       ATGGCATGTTGTGAGGGGACTTCGTGCTATGTGACGGAGTGTTAGTCTCACTCCG
MALCSPVDCAVQLCQSHS           244       ATGGCACTGTGTTCTCCGGTTGATTGCGCGGTGTAGTTGTGTCAGTCTCACTCCG
MAVCQSVDCALRWCQSHS           82        ATGGCAGTGTGTCAGAGTGTTGATTGCGCGCTGCGTTGGTGTTAGTCTCACTCCG
MAPCQSVRCSVQLCRSHS           100       ATGGCACCTTGTTAGAGTGTTCGTTGCAGTGTTTAGCTGTGTCGTTCTCACTCCG
```

MAACESLTCALIVCQSHS	434	ATGGCAGCTTGTGAGTCTCTGACTTGCGCTTTGATTGTTTGTTAGTCTCACTCCG
MAECVSLRCALPLCQSHS	79	ATGGCAGAGTGTGTTTCTCTGCGTTGCGCTCTTCCTCTTTGTTAGTCTCACTCCG
MAGCGAVACCLRSCQSHS	300	ATGGCAGGGTGTGGGGCTGTGGCTTGCTGTCTGCGTTCTTGTTAGTCTCACTCCG
MAQCRSIVCCLSDCKSHS	96	ATGGCATAGTGTAGGTCTATTGTGTGCTGTCTTTCGGATTGTAAGTCTCACTCCG
MATCVELQCKTLQCQSHS	81	ATGGCAACTTGTGTTGAGTTGTAGTGCAAGACTTTGTAGTGTTAGTCTCACTCCG
MAQCIGLSCESRECQSHS	79	ATGGCACAGTGTATTGGGTTGTCGTGCGAGAGTCGTGAGTGTTAGTCTCACTCCG
MAVCVGYGCSRIQCQSHS	115	ATGGCAGTTTGTGTGGGTTATGGGTGCAGTCGTATTCAGTGTTAGTCTCACTCCG
MAACVGMDCHWWRCQSHS	110	ATGGCAGCTTGTGTTGGGATGGATTGCCATTGGTGGCGTTGTTAGTCTCACTCCG
MAKCQDVTCSLVRCTSHS	91	ATGGCAAAGTGTTAGGATGTGACGTGCTCGTTGGTTCGTTGTACGTCTCACTCCG
MAQCSLVVCQLFRCPSHS	88	ATGGCATAGTGTTCTCTGGTTGTGTGCCAGCTGTTTAGGTGTCCTTCTCACTCCG
MAGCSSTTCQSRACGSHS	232	ATGGCAGGGTGTAGTAGTACTACGTGCCAGTCTCGTGCGTGTGGTTCTCACTCCG
MAACDLVDCQQRACGSHS	80	ATGGCAGCGTGTGATTTGGTGGATTGCTAGCAGCGTGCTTGTGGGTCTCACTCCG
MAECLIASCQYRACTSHS	96	ATGGCAGAGTGTCTGATTGCTTCGTGCTAGTATCGTGCGTGTACTTCTCACTCCG
MAECRLQRCHYVACSSHS	81	ATGGCAGAGTGTAGGTTGTAGCGTTGCCATTATGTTGCTTGTTCGTCTCACTCCG
MASCISRKCLAMQCHSHS	300	ATGGCATCGTGTATTAGTCGGAAGTGCTTGGCTATGTAGTGTCATTCTCACTCCG
MASCAAERCYVQHCMSHS	82	ATGGCATCTTGTGCTGCGGAGAGGTGCTATGTGTAGCATTGTATGTCTCACTCCG
MAYCVWDKCLLQLCGSHS	156	ATGGCATATTGTGTGTGGGATAAGTGCTTGTTGTAGTTGTGTGGGTCTCACTCCG
MAGCVSDKCIIQYCNSHS	76	ATGGCAGGGTGTGTTAGTGATAAGTGCATTATTTAGTATTGTAATTCTCACTCCG
MASCVVGKCLVQYCASHS	146	ATGGCATCTTGTGTTGTTGGGAAGTGCTTGGTTTAGTATTGTGCGTCTCACTCCG
MAACVMGKCLMQTCVSHS	103	ATGGCAGCTTGTGTGATGGGGAAGTGCTTGATGTAGACTTGTGTTTCTCACTCCG
MAPCVEGKCLKQSCSSHS	83	ATGGCACCGTGTGTGGAGGGTAAGTGCTTGAAGTAGTCGTGTAGTTCTCACTCCG
MAGCVLGKCLQDFCQSHS	87	ATGGCAGGTTGTGTTTTGGGGAAGTGCCTGTAGGATTTTTGTCAGTCTCACTCCG
MARCSSQVCRTTVCLSHS	189	ATGGCACGGTGTAGTAGTCAGGTGTGCAGGACGACTGTGTGTTTGTCTCACTCCG
MAQCARTVCRGSECLSHS	124	ATGGCACAGTGTGCGCGTACGGTTTGCCGTGGTTCGGAGTGTCTGTCTCACTCCG
Group5		
MAWCKVRECQRLGCVSHS	271	ATGGCATGGTGTAAGGTGCGGGAGTGCCAGAGGTTGGGTTGTGTTTCTCACTCCG
MAVCIGRSCQNLGCVSHS	104	ATGGCAGTTTGTATTGGGCGGTCTTGCTAGAATCTGGGGTGTGTTTCTCACTCCG
MALCQAHLCNPVSCRSHS	228	ATGGCACTTTGTTAGGCTCATTTGTGCAATCCGGTTTCTTGTCGGTCTCACTCCG
MALCVGCHCGPVSCQSHS	77	ATGGCATTGTGTGTGGGTTGTCATTGCGGTCCTGTTTCGTGTTAGTCTCACTCCG
MAHCSGSLCSQLSCVSHS	172	ATGGCACATTGTTCTGGGTCTTTGTGCTCGTAGCTTTCTTGTGTGTCTCACTCCG
MATCQIALCGQVLCVSHS	105	ATGGCAACGTGTTAGATTGCGCTGTGCGGTTAGGTTTTGTGTGTCTCACTCCG
MASCHQGLCESPHCQSHS	77	ATGGCAAGTTGTCATCAGGGGCTGTGCGAGTCTCCTCATTGTTAGTCTCACTCCG
MAVCLRQTCSSANCGSHS	425	ATGGCAGTGTGTCTGCGGTAGACGTGCAGTTCTGCGAATTGTGGTTCTCACTCCG
MAGCLPQACTVANCASHS	83	ATGGCAGGTTGTCTTCCTTAGGCGTGCACTGTGGCGAATTGTGCTTCTCACTCCG
MALCLHFICDQQLCASHS	142	ATGGCACTTTGTCTTCATTTTATTTGCGATCAGTAGCTGTGTGCGTCTCACTCCG
MAVCVLFVCNESLCSSHS	93	ATGGCAGTTTGTGTTCTGTTTGTGTGCAATGAGTCGCTGTGTTCTTCTCACTCCG
MAACVQMLCGVSLCASHS	93	ATGGCAGCTTGTGTTTAGATGCTTTGCGGGGTTTCGTTGTGTGCTTCTCACTCCG
MAICGQFLCFYSQCASHS	131	ATGGCAATTTGTGGGTAGTTTCTTTGCTTTTATTCGTAGTGTGCTTCTCACTCCG
Group6		
MAWCWGWGCFSSMCSSHS	96	ATGGCATGGTGTTGGGGGTGGGGGTGCTTTTCGAGTATGTGTAGTTCTCACTCCG
MADCWQPFCVYSHCTSHS	271	ATGGCAGATTGTTGGTAGCCGTTTTGCGTTTATTCGCATTGTACTTCTCACTCCG
MAICWAPYCTNLACTSHS	178	ATGGCAATTTGTTGGGCTCCGTATTGCACGAATTTGGCGTGTACGTCTCACTCCG
MAVCWLPQCISHLSHS	84	ATGGCAGTGTGTTGGTTGCCTTAGTGCATTTCTCATTTGTCTCACTCCG
Group7		
MAICDGVVCSFGKCGSHS	83	ATGGCAATTTGTGATGGGGTGGTGTGCTCTTTTGGGAAGTGTGGGTCTCACTCCG
MAFCQQVLCALGACSSHS	80	ATGGCATTTTGTCAGTAGGTGCTGTGCGCGCTGGGTGCGTGTTCTTCTCACTCCG
MAQCWQMPCSLGSCPSHS	344	ATGGCATAGTGTTGGCAGATGCCGTGCTCGTTGGGGTCGTGTCCTTCTCACTCCG
MAVCGCVGCQVGDCLSHS	150	ATGGCAGTGTGTGGTTGTGTTGGGTGCTAGGTTGGTGATTGTTTGTCTCACTCCG
MAVCHAIWCMVGHCLSHS	81	ATGGCAGTTTGTCATGCGATTTGGTGCATGGTGGGTCATTGTTTGTCTCACTCCG
MASCGWIGCLMGICYSHS	102	ATGGCAAGTTGTGGGTGGATTGGTTGCTTGATGGGGATTTGTTATTCTCACTCCG
MAMCGFSSCVDGICQSHS	102	ATGGCAATGTGTGGTTTTTCTTCGTGCGTTGATGGTATTTGTTAGTCTCACTCCG
MAMCGLPGCQLDRCYSHS	86	ATGGCAATGTGTGGGTTGCCGGGTTGCTAGTTGGATCGTTGTTATTCTCACTCCG
MAMCDDRGCGLGMCQSHS	138	ATGGCAATGTGTGATGATCGTGGTTGCGGTCTGGGGATGTGTTAGTCTCACTCCG
MACCGRGSCGLGQCQSHS	95	ATGGCATGTTGTGGGAGGGGGTCTTGCGGTCTGGGTCAGTGTTAGTCTCACTCCG
MAECRGPRCSGGLCGSHS	168	ATGGCAGAGTGTAGGGGGCCGAGGTGCTCGGGGGGTCTTTGTGGGTCTCACTCCG
MANCRGPGCQQGYCGSHS	81	ATGGCAAATTGTCGTGGTCCTGGTTGCCAGTAGGGGTATTGTGGGTCTCACTCCG
MATCERPGCRVGPCSSHS	87	ATGGCAACTTGTGAGCGTCCGGGTTGCAGGGTTGGTCCGTGTTCGTCTCACTCCG
Group8		
MARCLSPGCSQSLCCSHS	99	ATGGCACGGTGTCTTAGTCCGGGTTGCAGTCAGTCTTTGTGTTGTTCTCACTCCG
MARCAIRGCVKSLCCSHS	76	ATGGCAAGGTGTGCTATTCGGGGTTGCGTTAAGTCTTTGTGTTGTTCTCACTCCG
MAECLMPGCSSRLCQSHS	94	ATGGCAGAGTGTCTGATGCCTGGTTGCAGTTCGCGGCTTTGTTAGTCTCACTCCG
MAECLSRTCTQALCISHS	120	ATGGCAGAGTGTCTTAGTCGTACGTGCACTTAGGCTCTTTGTATTTCTCACTCCG
MAACLSDTCRQGLCSSHS	87	ATGGCAGCGTGTTTGTCGGATACGTGCCGTTAGGGGTTGTGTTCGTCTCACTCCG
Group9		
MAECGGIPCTQSTCVSHS	100	ATGGCAGAGTGTGGTGGGATTCCGTGCACTTAGTCTACTTGTGTTTCTCACTCCG
MAGCAGLPCSQPSCWSHS	79	ATGGCAGGTTGTGCGGGGCTTCCTTGCTCTTAGCCGTCGTGTTGGTCTCACTCCG
MAGCALAPCTQRYCASHS	87	ATGGCAGGGTGTGCGTTGGCGCCGTGCACGTAGCGTTATTGTGCTTCTCACTCCG
Group10		
MATCYSSWCFGWWCQSHS	84	ATGGCAACGTGTTATTCGTCGTGGTGCTTTGGGTGGTGGTGTTAGTCTCACTCCG
MAQCWADTCLRWVCLSHS	98	ATGGCATAGTGTTGGGCGGATACGTGCTTGCGTTGGGTGTGTCTTTCTCACTCCG
MAHCWVSACRHWRCQSHS	795	ATGGCACATTGTTGGGTGAGTGCTTGCCGGCATTGGCGGTGTTAGTCTCACTCCG
MAQCRPRLCVVWNCQSHS	133	ATGGCACAGTGTCGGCCTCGGTTGTGCGTGGTTTGGAATTGTTAGTCTCACTCCG

```
MAQCDTRLCPYWSCDSHS          77      ATGGCATAGTGTGATACGCGTTTGTGCCCTTATTGGAGTTGTGATTCTCACTCCG

 Group11
MALCSRSGCAYGCWSHS          108       ATGGCACTTTGTTCTCGTTCGGGTTGCGCGTATGGGTGTTGGTCTCACTCCG
MAMCCSSKCLLPNCWSHS          97       ATGGCAATGTGTTGTTCGTCTAAGTGCCTGTTGCCTAATTGTTGGTCTCACTCCG
MAGCQSFGCAGRTCWSHS         104       ATGGCAGGGTGTTAGAGTTTTGGGTGCGCTGGGCGTACTTGTTGGTCTCACTCCG
MAVCRFATCAGPQCWSHS         206       ATGGCAGTGTGTCGGTTTGCGACTTGTGCTGGGCCGTAGTGTTGGTCTCACTCCG
MAQCMVAGCALFECRSHS         124       ATGGCATAGTGTATGGTGGCTGGTTGCGCTTTGTTTGAGTGTCGGTCTCACTCCG

 Group12
MALCLMQPCASFACMSHS          94       ATGGCACTTTGTCTTATGTAGCCGTGCGCTTCGTTTGCGTGTATGTCTCACTCCG
MALCVSQHCDRFGCLSHS         105       ATGGCATTGTGTGTGTCTTAGCATTGCGATCGTTTTGGGTGTCTGTCTCACTCCG
MALCRSSVCSEGGCRSHS          87       ATGGCACTGTGTAGGAGTTCGGTGTGCTCTGAGGGTGGTTGTAGGTCTCACTCCG
MAVCPTFICTAMGCRSHS         279       ATGGCAGTTTGTCCTACTTTTATTTGCACGGCTATGGGGTGTAGGTCTCACTCCG

 Group13
MAQCGGRSCQGSSCPSHS          90      ATGGCACAGTGTGGTGGTAGGAGTTGCCAGGGGTCGTCGTGTCCGTCTCACTCCG
MAYCCGRDCQSETCQSHS         141      ATGGCATATTGTTGTGGTAGGGATTGCTAGTCTGAGACTTGTCAGTCTCACTCCG
MAACWCRQCQLHVCLSHS          96      ATGGCAGCTTGTTGGTGTCGGCAGTGCTAGCTTCATGTGTGTTTGTCTCACTCCG
MATCDLRSCQSDVCLSHS          84      ATGGCAACGTGTGATCTTCGGTCGTGCTAGTCTGATGTGTGTTTGTCTCACTCCG
MAGCVWRRCGLEACSSHS         136      ATGGCAGGTTGTGTGTGGCGGCGTTGCGGTTTGGAGGCTTGTTCGTCTCACTCCG
MAVCVMHKCQLNACSSHS         169      ATGGCAGTGTGTGTGATGCATAAGTGCTAGCTGAATGCGTGTTCTTCTCACTCCG

 Group14
MATCVRLYCMGLTCVSHS          81      ATGGCAACTTGTGTTCGTCTGTATTGCATGGGTCTTACGTGTGTGTCTCACTCCG
MAKCLRGSCGGLSCYSHS          81      ATGGCAAAGTGTTTGCGTGGTTCGTGCGGTGGGCTTTCTTGTTATTCTCACTCCG
MAVCNQRWCGGLSCRSHS         141      ATGGCAGTTTGTAATCAGCGTTGGTGCGGGGGGCTTAGTTGTCGGTCTCACTCCG
MAWCWYIGCAGIGCASHS         322      ATGGCATGGTGTTGGTATATTGGTTGCGCTGGTATTGGGTGTGCGTCTCACTCCG
MARCTLLGCNGVSCASHS         199      ATGGCACGGTGTACGCTTCTGGGGTGCAATGGTGTTTCGTGTGCTTCTCACTCCG
MASCDQIYCVPFSCASHS          86      ATGGCATCGTGTGATTAGATTTATTGCGTTCCGTTTTCTTGTGCGTCTCACTCCG
MACCDQMACEAICCVSHS         161      ATGGCATGTTGTGATTAGATGGCTTGCGAGGCTATTTGTTGTGTTTCTCACTCCG

 Group15
MAGCHSWQCTVIMCKSHS          90      ATGGCAGGTTGTCATTCTTGGTAGTGCACGGTTATTATGTGTAAGTCTCACTCCG
MAACQLWPCDVDNCVSHS          96      ATGGCAGCTTGTTAGTTGTGGCCTTGCGATGTGGATAATTGTGTGTCTCACTCCG
MALCRAWDCANCACTSHS         146      ATGGCACTGTGTAGGGCTTGGGATTGCGCTAATTGTGCTTGTACTTCTCACTCCG
MARCSKWSCQHMGCVSHS         167      ATGGCACGGTGTAGTAAGTGGTCGTGCCAGCATATGGGTTGTGTCTCTCACTCCG

 Group16
MAGCRAVTCAKQRCCSHS         103       ATGGCAGGGTGTCGTGCTGTGACTTGCGCTAAGTAGAGGTGTTGTTCTCACTCCG
MAGCQEQGCEQLRCCSHS         192       ATGGCAGGTTGTTAGGAGCAGGGGTGCGAGCAGTTGCGTTGTTGTTCTCACTCCG
MAGCGRNICVKNLCCSHS         114       ATGGCAGGGTGTGGGCGGAATATTTGCGTTAAGAATCTTTGTTGTTCTCACTCCG
MASCGSQNCWLRVCCSHS         126       ATGGCATCTTGTGGGTCGTAGAATTGCTGGCTGCGGGTGTGTTGTTCTCACTCCG
MASCQYQHCWFHVCCSHS         121       ATGGCATCGTGTCAGTATTAGCATTGCTGGTTTCATGTTTGTTGTTCTCACTCCG

 Group17
MALCSWGTCMQDLCVSHS          80      ATGGCACTTTGTAGTTGGGGTACTTGCATGTAGGATCTTTGTGTTTCTCACTCCG
MAVCLWGGCKQTSCSSHS          88      ATGGCAGTGTGTTTGTGGGGTGGGTGCAAGTAGACGAGTTGTTCTTCTCACTCCG
MAGCAWQICKEQTCASHS         101      ATGGCAGGTTGTGCGTGGCAGATTTGCAAGGAGTAGACTTGTGCGTCTCACTCCG
MAVCEWQLCVSDPCFSHS         101      ATGGCAGTTTGTGAGTGGTAGCTTTGCGTTTCGGATCCGTGTTTTTCTCACTCCG
MAICSWTWCAREGCDSHS         162      ATGGCAATTTGTTCGTGGACGTGGTGCGCTCGTGAGGGGTGTGATTCTCACTCCG

 Group18
MAVCATYPCPSLVCTSHS          89      ATGGCAGTTTGTGCTACGTATCCGTGCCCGTCTTTGGTGTGTACTTCTCACTCCG
MAGCGDYACPGLQCMSHS         160      ATGGCAGGTTGTGGGGATTATGCGTGCCCTGGGCTGTAGTGTATGTCTCACTCCG
MAGCWARPCPLALCQSHS         178      ATGGCAGGGTGTTGGGCTAGGCCGTGCCCTCTTGCGTTGTGTTAGTCTCACTCCG
MAACDARPCPQTYCLSHS         307      ATGGCAGCGTGTGATGCTCGTCCGTGCCCTTAGACTTATTGTTTGTCTCACTCCG

 Group19
MANCQLAVCQGSYCCSHS          78      ATGGCAAATTGTTAGCTTGCTGTTTGCCAGGGGTCTTATTGTTGTTCTCACTCCG
MALCTRSVCQVSYCISHS         170      ATGGCATTGTGTACGAGGTCGGTTTGCTAGGTGTCTTATTGTATTTCTCACTCCG
MAACQMNYCRSDYCKSHS         111      ATGGCAGCTTGTTAGATGAATTATTGCAGGTCGGATTATTGTAAGTCTCACTCCG

 Group MIXED
MAKCSDADCCGLRCGSHS         140       ATGGCAAAGTGTTCGGATGCGGATTGCTGTGGTTTGCGGTGTGGGTCTCACTCCG
MAKCSDQPCELMACKSHS         319       ATGGCAAAGTGTTCTGATTAGCCGTGCGAGTTGATGGCTTGTAAGTCTCACTCCG
MASCTLAECDCPACQSHS          89       ATGGCAAGTTGTACGCTTGCGGAGTGCGATTGTCCTGCTTGTTAGTCTCACTCCG
MAVCALISCVCYSCPSHS          92       ATGGCAGTGTGTGCGCTTATTTCTTGCGTGTGTTATTCTTGTCCTTCTCACTCCG
MANCVLGTCGRRPCVSHS         101       ATGGCAAATTGTGTGTTGGGTACTTGCGGTCGGCGGCCTTGTGTGTCTCACTCCG
MASCAVAHCGLSPCDSHS         186       ATGGCATCGTGTGCGGTTGCTCATTGCGGTTTGAGTCCGTGTGATTCTCACTCCG
MARCTKSLCCQSHS              81       ATGGCACGGTGTACTAAGTCGCTTTGCTGTCAGTCTCACTCCG
MAGCSKNLCCVRNSHS            79       ATGGCAGGTTGTTCGAAGAATTTGTGCTGTGTTCGTAATTCTCACTCCG
MACCRVGGCITLRCDSHS          90       ATGGCATGTTGTCGTGTTGGGGGGTGCATTACGCTGAGGTGTGATTCTCACTCCG
MAACQVAGCLWEHCDSHS          99       ATGGCAGCGTGTCAGGTGGCTGGTTGCTTGTGGGAGCATTGTGATTCTCACTCCG
MATCMRLGCCRDNCWSHS         121       ATGGCAACTTGTATGCGTCTGGGGTGCTGTCGGGATAATTGTTGGTCTCACTCCG
MARCLEQGCCSDECHSHS         153       ATGGCACGGTGTTTGGAGTAGGGGTGCTGTAGTGATGAGTGTCATTCTCACTCCG
MAICLLVCCKRIGCYSHS         107       ATGGCAATTTGTTTGCTTGTTTGTTGCAAGCGGATTGGTTGTTATTCTCACTCCG
MAVCVLLVCQFVCCYSHS          87       ATGGCAGTGTGTGTGTTGTTGGTGTGCCAGTTTGTGTGTTGTTATTCTCACTCCG
```

MASCYYGNCMQQPCYSHS	184	ATGGCATCTTGTTATTATGGGAATTGCATGCAGTAGCCGTGTTATTCTCACTCCG
MALCYRLTCQQLPCYSHS	80	ATGGCACTGTGTTATCGGCTTACGTGCTAGCAGCTGCCTTGTTATTCTCACTCCG
MARCALCLCGQGESLSHS	193	ATGGCACGTTGTGCGCTGTGTCTGTGCGGTTAGGGTGAGTCTTTGTCTCACTCCG
MAVCGPLQCIGWDCRSHS	112	ATGGCAGTGTGTGGGCCTTTGTAGTGCATTGGTTGGGATTGTCGGTCTCACTCCG
MAQCVPFKCVPGACDSHS	104	ATGGCATAGTGTGTTCCGTTTAAGTGCGTGCCGGGGGCGTGTGATTCTCACTCCG
MAQCFSRLCLRELCRSHS	84	ATGGCACAGTGTTTTTCGCGTTTGTGCCTTCGTGAGCTGTGTAGGTCTCACTCCG
MATCVSNRCCDWRCLSHS	80	ATGGCAACTTGTGTTTCGAATAGGTGCTGTGATTGGCGGTGTCTTTCTCACTCCG
MARCCALRCWLLSCPSHS	218	ATGGCACGGTGTTGTGCGCTTAGGTGCTGGTTGTTGTCTTGTCCTTCTCACTCCG
MARCCVIQCDHGACMSHS	93	ATGGCAAGGTGTTGTGTGATTTAGTGCGATCATGGGGCTTGTATGTCTCACTCCG
MAACGSRGCWQAGCYSHS	118	ATGGCAGCGTGTGGTAGTCGTGGGTGCTGGTAGGCTGGTTGTTATTCTCACTCCG
MASCGGDRCPWDGCWSHS	107	ATGGCATCTTGTGGTGGTGATCGTTGCCCTTGGGATGGTTGTTGGTCTCACTCCG
MAKCWDRACWGQFCTSHS	113	ATGGCAAAGTGTTGGGATCGTGCGTGCTGGGGGTAGTTTTGTACGTCTCACTCCG
MAYCQPAWCWCRNCSSHS	249	ATGGCATATTGTTAGCCTGCGTGGTGCTGGTGTCGTAATTGTTCGTCTCACTCCG
MAYCDVTVCWRGSCQSHS	78	ATGGCATATTGTGATGTTACGGTTTGCTGGCGGGGGAGTTGTTAGTCTCACTCCG
MANCSYGACWIGSCQSHS	126	ATGGCAAATTGTTCGTATGGGGCGTGCTGGATTGGGTCGTGTTAGTCTCACTCCG
MAMCVGQACWVRACRSHS	83	ATGGCAATGTGTGTTGGTCAGGCGTGCTGGGTGAGGGCTTGTCGGTCTCACTCCG
MAGCPGQVCWPVSCSSHS	167	ATGGCAGGGTGTCCGGGTTAGGTGTGCTGGCCTGTTTCTTGTAGTTCTCACTCCG
MASCMLAPCSLVICWSHS	91	ATGGCATCGTGTATGTTGGCTCCGTGCAGTTTGGTGATTTGTTGGTCTCACTCCG
MARCPLSHCAEQLCWSHS	86	ATGGCACGGTGTCCGTTGTCGCATTGCGCTGAGTAGTTGTGTTGGTCTCACTCCG
MAPCHDAQCARDQCDSHS	91	ATGGCACCGTGTCATGATGCTTAGTGCGCTCGGGATCAGTGTGATTCTCACTCCG
MALCQEGNCFAPTCASHS	251	ATGGCACTTTGTTAGGAGGGGAATTGCTTTGCTCCGACTTGTGCGTCTCACTCCG
MALCFDDVCMAGPCLSHS	83	ATGGCATTGTGTTTTGATGATGTTTGCATGGCGGGTCCGTGTTTGTCTCACTCCG
MAACRTSQCCPQGCASHS	297	ATGGCAGCGTGTCGGACGAGTCAGTGCTGTCCGTAGGGTTGTGCGTCTCACTCCG
MAVCRQSLCDSVLCQSHS	99	ATGGCAGTTTGTCGTCAGAGTCTTTGCGATAGTGTGTTGTGTCAGTCTCACTCCG
MALCPRQRCGTVHCASHS	180	ATGGCACTGTGTCCTCGGTAGAGGTGCGGTACGGTTCATTGTGCGTCTCACTCCG
MAHCSLQQCGLWFCGSHS	110	ATGGCACATTGTAGTTTGCAGTAGTGCGGGTTGTGGTTTTGTGGTTCTCACTCCG
MADCAQHCCGVATCRSHS	198	ATGGCAGATTGTGCGTAGCATTGTTGCGGTGTTGCGACTTGTCGGTCTCACTCCG
MARCGFNACKPTQCLSHS	104	ATGGCACGGTGTGGGTTTAATGCGTGCAAGCCGACTTAGTGTCTGTCTCACTCCG
MALCGGGWCQLQACGSHS	101	ATGGCACTTTGTGGTGGGGGGTGGTGCTAGCTGCAGGCTTGTGGCTCTCACTCCG
MAFCHGRNCAVLSCSSHS	80	ATGGCATTTTGTCATGGTAGGAATTGCGCTGTGCTGTCGTGTTCTTCTCACTCCG

PK – Libraries 3×3, 4×4

Group1		
MALCFDQCRVSCASHS	180	ATGGCACTTTGTTTTGATCAGTGCCGTGTTTCGTGTGCTTCTCACTCCG
MAMCFDSCRVNCTSHS	138	ATGGCAATGTGTTTTGATTCTTGCAGGGTTAATTGTACGTCTCACTCCG
MAVCYPRCRVSCSSHS	273	ATGGCAGTGTGTTATCCGCGGTGCCGGGTTTCTTGTTCTTCTCACTCCG
MATCFHFCRVSCGSHS	230	ATGGCAACTTGTTTTCATTTTTGCCGGGTTTCGTGTGGGTCTCACTCCG
MATCFHHCRVNCSSHS	95	ATGGCAACGTGTTTTCATCATTGCCGGGTGAATTGTTCTTCTCACTCCG
MATCFHRCRVTCISHS	93	ATGGCAACTTGTTTTCATCGGTGCCGTGTGACGTGTATTTCTCACTCCG
MARCFALCRVTCESHS	183	ATGGCACGGTGTTTTGCGTTGTGCAGGGTGACTTGTGAGTCTCACTCCG
MAKCFRLCRVTCHSHS	98	ATGGCAAAGTGTTTTCGTCTTTGCCGTGTGACGTGTCATTCTCACTCCG
MADCYQLCRVSCESHS	501	ATGGCAGATTGTTATTAGCTGTGCCGGGTTTCGTGTGAGTCTCACTCCG
MAVCYALCRVSCDSHS	149	ATGGCAGTTTGTTATGCGCTTTGCAGGGTGAGTTGTGATTCTCACTCCG
MATCFTMCRVSCDSHS	117	ATGGCAACTTGTTTTACGATGTGCCGTGTGAGTTGTGATTCTCACTCCG
MAICFLLCRVSCPSHS	257	ATGGCAATTTGTTTTCTTCTTTGCCGTGTGAGTTGTCCGTCTCACTCCG
MATCFLVCRVGCASHS	238	ATGGCAACTTGTTTTCTGGTGTGCAGGGTTGGTTGTGCTTCTCACTCCG
MAGCFEVCRVSCSSHS	140	ATGGCAGGGTGTTTTGAGGTTTGCCGTGTTTCGTGTTCTTCTCACTCCG
MASCFRECRVACPSHS	100	ATGGCAAGTTGTTTTCGGGAGTGCCGGGTTGCGTGTCCGTCTCACTCCG
MAVCFQECRVDCPSHS	100	ATGGCAGTGTGTTTTCAGGAGTGCCGGGTTGATTGTCCTTCTCACTCCG
MAQCFSRCRVGCPSHS	159	ATGGCACAGTGTTTTTCGAGGTGCCGTGTGGGTTGTCCTTCTCACTCCG
MAQCFERCRVNCDSHS	127	ATGGCATAGTGTTTTGAGAGGTGCCGTGTTAATTGTGATTCTCACTCCG
MAECFIRCRVNCDSHS	101	ATGGCAGAGTGTTTTATTCGGTGCAGGGTTAATTGTGATTCTCACTCCG
MATCLVVCRVNCDSHS	120	ATGGCAACTTGTCTGGTGGTTTGCCGTGTTAATTGTGATTCTCACTCCG
MADCFHQCRVGCDSHS	138	ATGGCAGATTGTTTTCATCAGTGCCGGGTTGGGTGTGATTCTCACTCCG
MASCFWKCRVGCDSHS	106	ATGGCATCTTGTTTTTGGAAGTGCAGGGTGGGGTGTGATTCTCACTCCG
Group2		
MAWCFPGCRVLCVSHS	164	ATGGCATGGTGTTTTCCGGGTTGCAGGGTGTTGTGTGTTTCTCACTCCG
MAWCWERCRVLCVSHS	113	ATGGCATGGTGTTGGGAGCGGTGCCGGGTTCTGTGTGTGTCTCACTCCG
MAWCFMQCKVVCYSHS	97	ATGGCATGGTGTTTTATGTAGTGCAAGGTGGTTTGTTATTCTCACTCCG
MAWCFDVCRVGCLSHS	161	ATGGCATGGTGTTTTGATGTGTGCAGGGTGGGGTGTCTGTCTCACTCCG
MAWCFTKCRTGCTSHS	124	ATGGCATGGTGTTTTACTAAGTGCCGTACTGGTTGTACTTCTCACTCCG
Group3		
MAGCWQHCRVLCWSHS	166	ATGGCAGGTTGTTGGTAGCATTGCCGGGTTCTTTGTTGGTCTCACTCCG
MASCWSYCRVYCWSHS	135	ATGGCATCGTGTTGGTCGTATTGCCGTGTGTATTGTTGGTCTCACTCCG
MARCWETCRVSCWSHS	227	ATGGCACGGTGTTGGGAGACTTGCCGGGTTTCGTGTTGGTCTCACTCCG
MASCWERCRVRCWSHS	164	ATGGCAAGTTGTTGGGAGCGTTGCAGGGTGAGGTGTTGGTCTCACTCCG
MARCWQLCRVGCNSHS	157	ATGGCACGGTGTTGGCAGCTTTGCCCGGGTTGGGTGTAATTCTCACTCCG
MAQCWQLCRVSCSSHS	108	ATGGCACAGTGTTGGTAGCTGTGCCGGGTGTCTTGTAGTTCTCACTCCG
MARCWLLCRVSCESHS	143	ATGGCACGGTGTTGGTTGTTGTGCCGGGTTAGTTGTGAGTCTCACTCCG
MALCWELCRVRCQSHS	157	ATGGCATTGTGTTGGGAGTTGTGCCGGGTGCGGTGTCAGTCTCACTCCG
MANCWRLCRVDCLSHS	105	ATGGCAAATTGTTGGCGTCTGTGCCCGGGTTGATTGTCTGTCTCACTCCG
MAGCWTLCRVRCHSHS	134	ATGGCAGGGTGTTGGACGTTGTGCCGTGTTAGGTGTCATTCTCACTCCG
MAGCWTLCRASCYSHS	113	ATGGCAGGTTGTTGGACTCTGTGCCGTGCTAGTTGTTATTCTCACTCCG

MAGCWTECRVNCLSHS	120	ATGGCAGGTTGTTGGACGGAGTGCCGTGTGAATTGTTTGTCTCACTCCG
MAACWSVCRVGCMSHS	134	ATGGCAGCGTGTTGGTCGGTTTGCCGTGTTGGTTGTATGTCTCACTCCG
MAGCWSLCRVGCSSHS	133	ATGGCAGGGTGTTGGTCTTTGTGCAGGGTGGGGTGTTCTTCTCACTCCG
MAACWVRCRVGCISHS	109	ATGGCAGCGTGTTGGGTGCGTTGCCGGGTGGGGTGTATTTCTCACTCCG
MAACWLLCRAGCYSHS	118	ATGGCAGCTTGTTGGTTGCTGTGCAGGGCGGGGTGTTATTCTCACTCCG
MAACYRLCRVGCYSHS	94	ATGGCAGCGTGTTATCGGTTGTGCCGGGTGGGTTGTTATTCTCACTCCG
MAVCFGLCRVGCLSHS	112	ATGGCAGTGTGTTTTGGTCTTTGCCGTGTGGGGTGTTTGTCTCACTCCG
MAKCWGLCRVGCLSHS	103	ATGGCAAAGTGTTGGGGGCTTTGCAGGGTTGGTTGTTTGTCTCACTCCG
MAVCWGVCRVACLSHS	207	ATGGCAGTGTGTTGGGGTGTGTGCCGTGTGGCTTGTCTTTCTCACTCCG
MAQCWRDCRVLCMSHS	91	ATGGCATAGTGTTGGAGGGATTGCCGTGTGTTGTGTATGTCTCACTCCG
MAQCWWSCRVNCTSHS	146	ATGGCATAGTGTTGGTGGAGTTGCAGGGTGAATTGTACCTCTCACTCCG
MAQCWYSCRVQCYSHS	116	ATGGCATAGTGTTGGTATTCTTGCCGGGTGCAGTGTTATTCTCACTCCG
Group4		
MAACFQMCRVNCLSHS	186	ATGGCAGCGTGTTTTCAGATGTGCCGGGTTAATTGTCTTTCTCACTCCG
MASCFQTCRVYCLSHS	156	ATGGCATCTTGTTTTTAGACTTGCAGGGTTTATTGTTTGTCTCACTCCG
MASCFNSCRVVCISHS	179	ATGGCAAGTTGTTTTAATAGTTGCCGGGTGGTTTGTATTTCTCACTCCG
MAACFDSCRVVCLSHS	97	ATGGCAGCGTGTTTTGATAGTTGCAGGGTTGTGTGTCTGTCTCACTCCG
MAACFTQCRVMCPSHS	201	ATGGCAGCTTGTTTTACTTAGTGCCGTGTTATGTGTCCTTCTCACTCCG
MAECFTQCRVMCLSHS	144	ATGGCAGAGTGTTTTACGTAGTGCCGTGTGATGTGTCTGTCTCACTCCG
MADCFQGCRVFCSSHS	124	ATGGCAGATTGTTTTTAGGGTTGCCGTGTTTTTTGTTCGTCTCACTCCG
MAYCFQGCRVLCYSHS	107	ATGGCATATTGTTTTCAGGGGTGCCGTGTGTTGTGTTATTCTCACTCCG
MARCFSSCRVLCNSHS	118	ATGGCAAGGTGTTTTAGTAGTTGCCGTGTTCTGTGTAATTCTCACTCCG
MARCFSGCRVLCFSHS	91	ATGGCACGTTGTTTTTCGGGGTGCAGGGTGCTTTGTTTTTCTCACTCCG
MAQCFTGCRVSCGSHS	184	ATGGCATAGTGTTTTACTGGTTGCCGTGTTTCGTGTGGGTCTCACTCCG
MAQCFIGCRVNCGSHS	178	ATGGCATAGTGTTTTATTGGGTGCAGGGTTAATTGTGGGTCTCACTCCG
MAPCFRGCRVNCSSHS	142	ATGGCACCGTGTTTTAGGGGGTGCAGGGTTAATTGTTCGTCTCACTCCG
MASCFTGCRVACPSHS	137	ATGGCATCGTGTTTTACGGGTTGCCGTGTGGCTTGTCCGTCTCACTCCG
MASCFSGCRVACRSHS	134	ATGGCATCGTGTTTTAGTGGGTGCCGTGTGGCTTGTCGGTCTCACTCCG
MAKCFQGCRVHCVSHS	155	ATGGCAAAGTGTTTTCAGGGGTGCCGGGTGCATTGTGTGTCTCACTCCG
MAKCWQGCRVNCMSHS	145	ATGGCAAAGTGTTGGTAGGGGTGCAGGGTTAATTGTATGTCTCACTCCG
MAECLHGCRVACLSHS	196	ATGGCAGAGTGTTTGCATGGTTGCCGGGTGGCGTGTCTGTCTCACTCCG
MAQCFQLCRTACVSHS	137	ATGGCACAGTGTTTTTAGCTTTGCAGGACGGCTTGTGTGTCTCACTCCG
MAQCFRSCRVACVSHS	134	ATGGCACAGTGTTTTCGGTCTTGCAGGGTGGCGTGTGTGTCTCACTCCG
Group5		
MAKCFQACRTLCFSHS	178	ATGGCAAAGTGTTTTTAGGCTTGCAGGACGCTTTGTTTTTCTCACTCCG
MAKCFMSCRGLCFSHS	99	ATGGCAAAGTGTTTTATGTCGTGCCGTGGGCTGTGTTTTTCTCACTCCG
MAFCFQNCKALCFSHS	202	ATGGCATTTTGTTTTCAGAATTGCAAGGCTCTTTGTTTTTCTCACTCCG
MALCFANCRALCFSHS	127	ATGGCACTTTGTTTTGCTAATTGCCGGGCGTTGTGTTTTTCTCACTCCG
MATCFQTCRALCWSHS	115	ATGGCAACTTGTTTTCAGACTTGCAGGGCTCTGTGTTGGTCTCACTCCG
MATCFQVCRALCVSHS	105	ATGGCAACTTGTTTTCAGGTGTGCCGTGCGCTGTGTGTTTCTCACTCCG
MAPCFQPCRSICMSHS	92	ATGGCACCGTGTTTTTAGCCTTGCAGGAGTATTTGTATGTCTCACTCCG
MAECFRLCRTLCSSHS	166	ATGGCAGAGTGTTTTCGTCTGTGCCGTACTCTTTGTTCTTCTCACTCCG
MAHCFRPCRALCPSHS	102	ATGGCACATTGTTTTCGTCCGTGCCGGGCGTTGTGTCCGTCTCACTCCG
Group6		
MAGCFTPCRTMCWSHS	317	ATGGCAGGGTGTTTTACTCCGTGCAGGACTATGTGTTGGTCTCACTCCG
MAKCFPKCRVMCWSHS	187	ATGGCAAAGTGTTTTCCGAAGTGCCGTGTGATGTGTTGGTCTCACTCCG
MARCFVSCRGVCWSHS	92	ATGGCACGGTGTTTTGTGAGTTGCAGGGGTGTTTGTTGGTCTCACTCCG
MAECFLVCRVACWSHS	108	ATGGCAGAGTGTTTTTTGGTGTGCAGGGTTGCGTGTTGGTCTCACTCCG
MASCLWVCRVACWSHS	91	ATGGCAAGTTGTCTGTGGGTGTGCAGGGTGGCGTGTTGGTCTCACTCCG
MAQCFWFCRTGCWSHS	169	ATGGCATAGTGTTTTTGGTTTTGCAGGACTGGGTGTTGGTCTCACTCCG
MARCFSLCRVQCWSHS	120	ATGGCAAGGTGTTTTTCGTTGTGCCGGGTTCAGTGTTGGTCTCACTCCG
MAHCYSVCRVNCWSHS	117	ATGGCACATTGTTATAGTGTTTGCCGTGTGAATTGTTGGTCTCACTCCG
Group7		
MARCYMPCRVNCVSHS	118	ATGGCACGTTGTTATATGCCTTGCCGTGTTAATTGTGTTTCTCACTCCG
MALCFVPCRVDCVSHS	109	ATGGCATTGTGTTTTTGTGCCGTGCCGGGTGGATTGTGTTTCTCACTCCG
MAQCFSPCRVECLSHS	101	ATGGCATAGTGTTTTTCTCCGTGCCGTGTTGAGTGTCTTTCTCACTCCG
Group8		
MANCFQTCRVSCYSHS	350	ATGGCAAATTGTTTTCAGACGTGCCGTGTGTCTTGTTATTCTCACTCCG
MAKCFQACRASCYSHS	137	ATGGCAAAGTGTTTTTAGGCTTGCCGTGCTTCGTGTTATTCTCACTCCG
MASCFDRCRVGCYSHS	241	ATGGCATCTTGTTTTGATCGGTGCCGTGTTGGTTGTTATTCTCACTCCG
MAECFDRCRVSCYSHS	95	ATGGCAGAGTGTTTTGATAGGTGCAGGGTGAGTTGTTATTCTCACTCCG
MATCFRSCKVACYSHS	219	ATGGCAACTTGTTTTTCGTAGTTGCAAGGTTGCTTGTTATTCTCACTCCG
MADCFSSCKVACYSHS	195	ATGGCAGATTGTTTTAGTTCTTGCAAGGTGGCGTGTTATTCTCACTCCG
MAVCFDSCRAACYSHS	248	ATGGCAGTTTGTTTTGATTCGTGCAGGGCTGCGTGTTATTCTCACTCCG
Group9		
MALCFVACRVHCLSHS	175	ATGGCACTGTGTTTTGTTGCTTGCCGGGTGCATTGTCTGTCTCACTCCG
MALCFVGCRVMCISHS	142	ATGGCATTGTGTTTTGTGGGTTGCCGTGTGATGTGTATTTCTCACTCCG
MALCYTACRVLCASHS	153	ATGGCATTGTGTTATACGGCGTGCCGTGTTTTGTGTGCGTCTCACTCCG
Group10		
MAFCWSPCRVSCGSHS	232	ATGGCATTTTGTTGGTCTCCTTGCCGGGTGTCTTGTGGTTCTCACTCCG
MAFCWSQCRVSCMSHS	198	ATGGCATTTTGTTGGTCGTAGTGCCGGGTTTCGTGTATGTCTCACTCCG

```
MATCWTSCRVSCGSHS          97       ATGGCAACGTGTTGGACTAGTTGCAGGGTTTCTTGTGGGTCTCACTCCG

 Group11
MAKCYRQCMQNCRSHS         285        ATGGCAAAGTGTTATCGGCAGTGCATGCAGAATTGTCGGTCTCACTCCG
MAECYRSCHAMCRSHS          99       ATGGCAGAGTGTTATCGGAGTTGCCATGCTATGTGTAGGTCTCACTCCG
MAGCYQQCPMYCRSHS         130        ATGGCAGGGTGTTATTAGCAGTGCCCGATGTATTGTAGGTCTCACTCCG
MANCFVNCRVACHSHS         123        ATGGCAAATTGTTTTGTTAATTGCAGGGTTGCGTGTCATTCTCACTCCG
MADCFVRCRVACASHS         105        ATGGCAGATTGTTTTGTGCGTTGCCCGGGTTGCGTGTGCGTCTCACTCCG
MANCFLMCRAACISHS         134        ATGGCAAATTGTTTTCTGATGTGCCGGGCTGCTTGTATTTCTCACTCCG
MASCFVRCRTICYSHS         164        ATGGCAAGTTGTTTTGTTAGGTGCAGGACGATTTGTTATTCTCACTCCG
MASCFVQCRTQCYSHS         108        ATGGCATCTTGTTTTGTTCAGTGCCGGACGTAGTGTTATTCTCACTCCG
MARCFIYCRTLCMSHS         415        ATGGCAAGGTGTTTTATTTATTGCCCGGACGTTGTGTATGTCTCACTCCG
MAFCYVRCRVNCMSHS         206        ATGGCATTTTGTTATGTTCGTTGCCGTGTTAATTGTATGTCTCACTCCG
MAGCYTRCRVDCFSHS         249        ATGGCAGGTTGTTATACTCGGTGCCGGGTTGATTGTTTTTCTCACTCCG
MANCFMPCRVHCRSHS         113        ATGGCAAATTGTTTTATGCCTTGCAGGGTGCATTGTAGGTCTCACTCCG
MARCFGHCRVACESHS         123        ATGGCAAGGTGTTTTGGGCATTGCCCGGGTGGCTTGTGAGTCTCACTCCG
MARCFPLCRANCYSHS         136        ATGGCACGGTGTTTTCCTTTGTGCCGTGCTAATTGTTATTCTCACTCCG
MAHCLRLCRVACYSHS         227        ATGGCACATTGTTTGCGTCTGTGCAGGGTGGCGTGTTATTCTCACTCCG
MAVCFPLCRVPCISHS         278        ATGGCAGTGTGTTTTCCGTTGTGCAGGGTGCCGTGTATTTCTCACTCCG
MAVCYELCRVMCMSHS         107        ATGGCAGTTTGTTATGAGTTGTGCAGGGTTATGTGTATGTCTCACTCCG
MANCWYVCRALCVSHS         221        ATGGCAAATTGTTGGTATGTGTGCCGGGCGTTGTGTGTGTCTCACTCCG
MAVCWVSCRALCVSHS         192        ATGGCAGTGTGTTGGGTTTCGTGCCGGGCGCTGTGTGTGTCTCACTCCG
MAQCWIACRVVCLSHS         292        ATGGCACAGTGTTGGATTGCGTGCCGGGTTGTTTGTCTTTCTCACTCCG
MASCWFLCRGACYSHS         147        ATGGCAAGTTGTTGGTTTCTGTGCCGTGGGGCGTGTTATTCTCACTCCG
MAFCWFTCRGMCYSHS          97        ATGGCATTTTGTTGGTTTACTTGCCGGGGTATGTGTTATTCTCACTCCG
MAACWDRCRVACQSHS         211        ATGGCAGCTTGTTGGGATCGGTGCCGTGTTGCGTGTCAGTCTCACTCCG
MAVCWRPCRVACWSHS          95        ATGGCAGTTTGTTGGCGGCCTTGCAGGGTGGCTTGTTGGTCTCACTCCG
MAMCHTGCGVMCLSHS         169        ATGGCAATGTGTCATACGGGTTGCGGTGTTATGTGTCTTTCTCACTCCG
MAWCHTSCQVACSSHS         290        ATGGCATGGTGTCATACGAGTTGCCAGGTGGCTTGTTCTTCTCACTCCG
MAGCYSFCCILCISHS          94        ATGGCAGGTTGTTATTCTTTTTGCTGTATTCTTTGTATTTCTCACTCCG
MAVCFTGCCLLCPSHS         212        ATGGCAGTTTGTTTTACTGGGTGCTGTCTTTTGTGTCCGTCTCACTCCG

 Group12
MALCWERRCVMQSCTSHS        92         ATGGCACTGTGTTGGGAGCGTAGGTGCGTGATGTAGTCTTGTACGTCTCACTCCG
MASCYEKRCVIRTCRSHS        129        ATGGCAAGTTGTTATGAGAAGCGGTGCGTTATTCGTACGTGTCGTTCTCACTCCG
MAMCEQRACTFRECWSHS        417        ATGGCAATGTGTGAGTAGAGGGCGTGCACTTTTCGTGAGTGTTGGTCTCACTCCG
MATCERRACVMRECASHS        165        ATGGCAACTTGTGAGAGGCGTGCGTGCGTTATGCGTGAGTGTGCGTCTCACTCCG
MATCEKRCCPRCSSHS          114        ATGGCAACGTGTGAGAAGCGGTGCTGTCCGCGTTGTTCGTCTCACTCCG
MAICTKRCFNSCTSHS          151        ATGGCAATTTGTACTAAGCGGTGCTTTAATTCGTGTACGTCTCACTCCG

 Group13
MAKCKTRSCTSLICSSHS         113         ATGGCAAAGTGTAAGACTCGTTCGTGCACTTCTTTGATTTGTTCGTCTCACTCCG
MARCVTGRCSMMVCASHS         139         ATGGCAAGGTGTGTGACTGGGCGGTGCTCTATGATGGTTTGTGCTTCTCACTCCG
MARCASGLCIGFICLSHS         127         ATGGCACGTTGTGCGAGTGGTCTGTGCATTGGGTTTATTTGTCTGTCTCACTCCG

 Group14
MAQCSMLCAGDCASHS          92         ATGGCACAGTGTTCGATGTTGTGCGCGGGTGATTGTGCTTCTCACTCCG
MASCTLFCQASCVSHS          200        ATGGCAAGTTGTACGTTGTTTTGCCAGGCGAGTTGTGTGTCTCACTCCG
MASCGIICEGYCFSHS          146        ATGGCATCTTGTGGTATTATTTGCGAGGGTTATTGTTTTTCTCACTCCG
MAACLLVCSGCCGSHS          104        ATGGCAGCTTGTCTGCTTGTGTGCTCGGGTTGTTGTGGTTCTCACTCCG
MARCNLRCSDGCLSHS          114        ATGGCACGTTGTAATTTGCGGTGCTCGGATGGGTGTCTTTCTCACTCCG

 Group15
MAGCGRLCSEMCGSHS          123        ATGGCAGGGTGTGGGCGGTTGTGCTCGGAGATGTGTGGTTCTCACTCCG
MAGCGRACALGCLSHS          183        ATGGCAGGTTGTGGGAGGGCGTGCGCTCTGGGTTGTTTGTCTCACTCCG
MAMCGRNACSPLSCESHS        114        ATGGCAATGTGTGGTCGGAATGCTTGCTCGCCGCTGTCTTGTGAGTCTCACTCCG

 Group MIXED
MAICVSRSCLELWCLSHS        129        ATGGCAATTTGTGTGAGTCGTTCGTGCTTGGAGCTGTGGTGTTTGTCTCACTCCG
MAVCSCRCAAWCLSHS          283        ATGGCAGTTTGTTCGTGTAGGTGCGCTGCGTGGTGTCTTTCTCACTCCG
MAWCATDLCDACVCQSHS        309        ATGGCATGGTGTGCGACGGATCTGTGCGATGCTTGTGTGTGTTAGTCTCACTCCG
MAACGEIYCRVSHS            214        ATGGCAGCTTGTGGGGAGATTTATTGCCGTGTTTCTCACTCCG
MAYCLHCRVSHS              217        ATGGCATATTGTTTACATTGCCGTGTTTCTCACTCCG
MALCLGQRCESVVCLSHS        198        ATGGCACTTTGTTTGGGTCAGCGTTGCGAGTCTGTGGTGTGTCTTTCTCACTCCG
MARCLGHRCGAWICSSHS        287        ATGGCACGTTGTTTGGGTCATCGGTGCGGGGCGTGGATTTGTTCGTCTCACTCCG
MAACGVSCATRCGSHS          146        ATGGCAGCTTGTGGGGTTAGTTGCGCGACTAGGTGTGGTTCTCACTCCG
MAECHVRCGRHCGSHS          225        ATGGCAGAGTGTCATGTGCGGTGCGGTCGGCATTGTGGTTCTCACTCCG
MALCSSQMCSTRPCRSHS        110        ATGGCACTTTGTTCGAGTTAGATGTGCTCTACTCGTCCGTGTAGGTCTCACTCCG
MAMCNARICRLSPCSSHS        114        ATGGCAATGTGTAATGCGCGGATTTGCCGGTTGTCTCCGTGTAGTTCTCACTCCG
MAGCEGAWCALARCWSHS         95        ATGGCAGGTTGTGAGGGTGCGTGGTGCGCGCTTGCTCGGTGTTGGTCTCACTCCG
MAACWIGWCSLARCISHS        104        ATGGCAGCTTGTTGGATTGGTTGGTGCTCTCTGGCTAGGTGTATTTCTCACTCCG
MATCFMGCSNSCLSHS          125        ATGGCAACTTGTTTTATGGGGTGCTCTAATTCGTGTCTTTCTCACTCCG
MALCVFSCARSCLSHS          170        ATGGCACTGTGTGTTTTTTCTTGCGCTCGGTCTTGTTTGTCTCACTCCG
MAKCEIPCAVSCLSHS          114        ATGGCAAAGTGTGAGATTCCTTGCGCGGTTTCGTGTTTGTCTCACTCCG
MAQCENPCPTSCCSHS          159        ATGGCATAGTGTGAGAATCCGTGCCCTACGAGTTGTTGTTCTCACTCCG
MAYTSHS                   128        ATGGCATATACGTCTCACTCCG
MAFCSVWVREAYLCASHS         99        ATGGCATTTTGTTCGGTTTGGGTGCGTGAGGCGTATTTGTGTGCTTCTCACTCCG
MAGCLCFRCDGHVCDSHS        149        ATGGCAGGGTGTTTGTGTTTTAGGTGCGATGGTCATGTGTGTGATTCTCACTCCG
```

MAVCYRVWCLDWLCRSHS	125	ATGGCAGTTTGTTATCGGGTTTGGTGCTTGGATTGGCTTTGTCGTTCTCACTCCG
MAYCVWANCFGWMCLSHS	149	ATGGCATATTGTGTTTGGGCGAATTGCTTTGGTTGGATGTGTTTGTCTCACTCCG
MALCLWRCIWACPSHS	92	ATGGCACTTTGTCTGTGGCGGTGCATTTGGGCGTGTCCTTCTCACTCCG
MALCAIAQCGQHYCGSHS	117	ATGGCACTTTGTGCGATTGCGCAGTGCGGGTAGCATTATTGTGGTTCTCACTCCG
MAWCVSSCFQGCISHS	133	ATGGCATGGTGTGTTTCGAGTTGCTTTTAGGGTTGTATTTCTCACTCCG
MASCIQSCGYHCVSHS	141	ATGGCATCGTGTATTTAGTCGTGCGGTTATCATTGTGTGTCTCACTCCG
MAYCLDKCGSRCESHS	100	ATGGCATATTGTCTTGATAAGTGCGGGTCGAGGTGTGAGTCTCACTCCG
MAPCRFRCGPVCWSHS	185	ATGGCACCTTGTCGTTTTCGGTGCGGTCCGGTGTGTTGGTCTCACTCCG
MARCVVCCDMTCWSHS	199	ATGGCAAGGTGTGTGGTTTGTTGCGATATGACGTGTTGGTCTCACTCCG
MALCKVCCRPLCMSHS	93	ATGGCACTTTGTAAGGTGTGTTGCCGGCCGTTGTGTATGTCTCACTCCG
MAGCNDLCHSGCCSHS	120	ATGGCAGGTTGTAATGATTTGTGCCATTCGGGTTGTTGTTCTCACTCCG
MASCVDYCVGTCCSHS	105	ATGGCATCTTGTGTGGATTATTGCGTTGGTACGTGTTGTTCTCACTCCG
MACCKLLCSQVCRSHS	96	ATGGCATGTTGTAAGCTTCTTTGCTCTCAGGTGTGTAGGTCTCACTCCG
MAGCQTKCNSPCKSHS	129	ATGGCAGGGTGTTAGACTAAGTGCAATAGTCCTTGTAAGTCTCACTCCG
MAVCSLHGCWRIGCMSHS	282	ATGGCAGTTTGTTCTCTGCATGGTTGCTGGCGTATTGGTTGTATGTCTCACTCCG
MAPCCSQDCRVDYCASHS	127	ATGGCACCTTGTTGTTCGTAGGATTGCAGGGTGGATTATTGTGCGTCTCACTCCG
MAGCGFHDCQTRRCPSHS	103	ATGGCAGGGTGTGGTTTTCATGATTGCTAGACGCGTCGGTGTCCGTCTCACTCCG
MATCSWGKCQVTDCGSHS	295	ATGGCAACTTGTAGTTGGGGTAAGTGCTAGGTGACGGATTGTGGGTCTCACTCCG
MARCGCQGCSVADCSSHS	149	ATGGCAAGGTGTGGGTGTCAGGGGTGCAGTGTTGCTGATTGTAGTTCTCACTCCG
MAACRSWSCTCGLCYSHS	255	ATGGCAGCTTGTCGGTCGTGGTCGTGCACTTGTGGGTTGTGTTATTCTCACTCCG
MAKCRQPVCLPSFCASHS	185	ATGGCAAAGTGTAGGTAGCCTGTTTGCCTGCCGAGTTTTTGTGCTTCTCACTCCG
MALCVSGVCHHEACPSHS	132	ATGGCACTTTGTGTTAGTGGTGTGTGCCATCATGAGGCGTGTCCTTCTCACTCCG
MADCWDSVCYRVLCWSHS	704	ATGGCAGATTGTTGGGATAGTGTGTGCTATAGGGTGTTGTGTTGGTCTCACTCCG
MALCIGSVCQEADCWSHS	92	ATGGCATTGTGTATTGGTTCGGTGTGCTAGGAGGCTGATTGTTGGTCTCACTCCG
MATCHQLLCMAMSCVSHS	117	ATGGCAACTTGTCATTAGCTGTTGTGCATGGCGATGTCTTGTGTTTCTCACTCCG
MAQCGGILCGALPCFSHS	127	ATGGCATAGTGTGGTGGTATTCTTTGCGGGGCGCTTCCTTGTTTTTCTCACTCCG
MALCLSVMCTALECFSHS	114	ATGGCATTGTGTCTTAGTGTGATGTGCACTGCTTTGGAGTGTTTTTCTCACTCCG
MATCMSRLCSGGMCKSHS	201	ATGGCAACGTGTATGTCTCGTTTGTGCTCTGGTGGTATGTGTAAGTCTCACTCCG
MAMCDLSLCIGSLCASHS	244	ATGGCAATGTGTGATTTGAGTCTGTGCATTGGTTCTCTTTGTGCGTCTCACTCCG
MAVCTLDLCTGDGCGSHS	99	ATGGCAGTGTGTACGTTGGATTTGTGCACTGGTGATGGGTGTGGTTCTCACTCCG
MARCSDCMCASCRCQSHS	91	ATGGCACGTTGTTCGGATTGTATGTGCGCTTCGTGTCGTTGTTAGTCTCACTCCG
MAECDGGPCVQQVCKSHS	146	ATGGCAGAGTGTGATGGTGGGCCGTGCGTTCAGTAGGTTTGTAAGTCTCACTCCG

SA – Libraries 3×3, 4×4

Group1

MAQCQTRCTAVCASHS	702	ATGGCACAGTGTTAGACGAGGTGCACGGCTGTGTGTGCGTCTCACTCCG
MATCQTLCSRTCISHS	102	ATGGCAACTTGTTAGACTCTGTGCTCGCGTACTTGTATTTCTCACTCCG
MAGCQQRCGLVCCSHS	243	ATGGCAGGGTGTCAGCAGAGGTGCGGGCTGGTTTGTTGTTCTCACTCCG
MASCHTECQYICWSHS	198	ATGGCAAGTTGTCATACTGAGTGCTAGTATATTTGTTGGTCTCACTCCG
MASCQRACAWHCVSHS	471	ATGGCATCTTGTTAGAGGGCGTGCGCTTGGCATTGTGTGTCTCACTCCG
MATCQVTCVDHCLSHS	167	ATGGCAACTTGTTAGGTTACGTGCGTTGATCATTGTCTGTCTCACTCCG
MASCRDVCCSHCLSHS	795	ATGGCATCGTGTAGGGATGTGTGCTGTAGTCATTGTCTGTCTCACTCCG
MASCRPLCLSRCLSHS	184	ATGGCAAGTTGTAGGCCGTTGTGCCTTAGTCGTTGTCTGTCTCACTCCG
MARCSMTCQQCCESHS	273	ATGGCACGGTGTTCTATGACGTGCCAGCAGTGTTGTGAGTCTCACTCCG
MARCSFECSQSCLSHS	202	ATGGCACGTTGTTCTTTTGAGTGCAGTTAGAGTTGTTTGTCTCACTCCG
MARCAVRCGVPCLSHS	140	ATGGCAAGGTGTGCTGTGCGTTGCGGGGTTCCGTGTTTGTCTCACTCCG
MAQCEIFCRVKCISHS	199	ATGGCATAGTGTGAGATTTTTTGCAGGGTGAAGTGTATTTCTCACTCCG
MAQCRQECRPQCMSHS	129	ATGGCACAGTGTCGGTAGGAGTGCCGGCCGCAGTGTATGTCTCACTCCG
MARCSQVCRATCASHS	358	ATGGCACGTTGTTCTTAGGTGTGCAGGGCTACGTGTGCGTCTCACTCCG
MARCGEFCRVLCSSHS	111	ATGGCAAGGTGTGGGGAGTTTTGCCGTGTGCTTTGTTCTTCTCACTCCG
MAQCNRVCDVRCGSHS	123	ATGGCATAGTGTAATCGTGTGTGCGATGTTAGGTGTGGGTCTCACTCCG
MAQCGMACGLRCGSHS	121	ATGGCACAGTGTGGTATGGCGTGCGGGTTGAGGTGTGGTTCTCACTCCG
MAECVTDCQMRCGSHS	182	ATGGCAGAGTGTGTGACTGATTGCTAGATGCGGTGTGGTTCTCACTCCG
MARCVPSCSRHCSSHS	240	ATGGCACGGTGTGTGCCTAGTTGCAGTCGGCATTGTTCTTCTCACTCCG
MAQCAPMCSVRCSSHS	152	ATGGCACAGTGTGCTCCGATGTGCTCTGTTCGGTGTAGTTCTCACTCCG
MAECMSPCLRGCASHS	185	ATGGCAGAGTGTATGTCTCCTTGCTTGCGTGGTTGTGCTTCTCACTCCG
MAACVSYCVDGCWSHS	101	ATGGCAGCGTGTGTTAGTTATTGCGTGGATGGGTGTTGGTCTCACTCCG
MAYCSGACSSGCFSHS	143	ATGGCATATTGTTCTGGGGCGTGCAGTTCTGGTTGTTTTTCTCACTCCG
MARCASSCTTGCLSHS	114	ATGGCACGGTGTGCGTCGAGTTGCACTACGGGGTGTCTGTCTCACTCCG
MAHCKMLCTGGCYSHS	111	ATGGCACATTGTAAGATGTTGTGCACGGGTGGTTGTTATTCTCACTCCG
MAMCLMVCSVRCRSHS	216	ATGGCAATGTGTCTGATGGTTTGCAGTGTGAGGTGTCGGTCTCACTCCG
MAGCLQTCQASCRSHS	166	ATGGCAGGGTGTTTGTAGACTTGCCAGGCTTCTTGTCGTTCTCACTCCG
MAGCCLDCTLACRSHS	239	ATGGCAGGTTGTTGTCTGGATTGCACTTTGGCTTGTCGGTCTCACTCCG
MAGCVADCGLGCESHS	104	ATGGCAGGGTGTGTTGCGGATTGCGGGCTGGGGTGTGAGTCTCACTCCG
MAPCMRNCYQACRSHS	213	ATGGCACCTTGTATGAGGAATTGCTATTAGGCGTGTAGGTCTCACTCCG
MAVCGGVCGHACRSHS	195	ATGGCAGTGTGTGGGGGGGTGTGCGGTCATGCGTGTCGGTCTCACTCCG
MARCEIGCGHACRSHS	110	ATGGCACGGTGTGAGATTGGGTGCGGTCATGCTTGTCGTTCTCACTCCG
MASCHQSCGTRCMSHS	270	ATGGCATCTTGTCATCAGTCGTGCGGGACGCGGTGTATGTCTCACTCCG
MASCGESCGRSCKSHS	142	ATGGCATCGTGTGGGGAGTCGTGCGGTCGTTCTTGTAAGTCTCACTCCG
MAGCTQSCPCRCDSHS	447	ATGGCAGGGTGTACGTAGAGTTGCCCGTGTAGGTGTGATTCTCACTCCG
MAGCDSLCPERCQSHS	210	ATGGCAGGTTGTGATTCGTTGTGCCCGGAGCGGTGTTAGTCTCACTCCG
MAGCNLLCAYSCPSHS	110	ATGGCAGGTTGTAATCTTCTGTGCGCTTATTCTTGTCCTTCTCACTCCG

Group2

```
MASCVVSGCHPQVCPSHS          124       ATGGCAAGTTGTGTGGTTTCGGGGTGCCATCCGTAGGTGTGTCCGTCTCACTCCG
MAQCWKTGCHPQQCPSHS          116       ATGGCACAGTGTTGGAAGACTGGTTGCCATCCGTAGCAGTGTCCGTCTCACTCCG
MAPCQFQLCHPQVCPSHS          354       ATGGCACCGTGTCAGTTTTAGTTGTGCCATCCTCAGGTGTGTCCGTCTCACTCCG
MASCDPAHCHPQVCYSHS          555       ATGGCATCTTGTGATCCGGCGCATTGCCATCCTTAGGTGTGTTATTCTCACTCCG
MAWCSLSACHPQGCFSHS          180       ATGGCATGGTGTAGTCTGTCGGCGTGCCATCCTTAGGGGTGTTTTTCTCACTCCG
MAVCVMNGCHPQGCVSHS          161       ATGGCAGTTTGTGTGATGAATGGGTGCCATCCTTAGGGGTGTGTGTCTCACTCCG
MAYCHAYACHPQNCRSHS          483       ATGGCATATTGTCATGCTTATGCGTGCCATCCGTAGAATTGTCGTTCTCACTCCG
MAKCLSFACHPQNCTSHS          260       ATGGCAAAGTGTCTGTCTTTTGCTTGCCATCCTCAGAATTGTACTTCTCACTCCG
MAFCGWLGCHPQGCSSHS          490       ATGGCATTTTGTGGGTGGCTGGGGTGCCATCCGTAGGGTTGTTCGTCTCACTCCG
MAVCGELSCHPQFCGSHS          112       ATGGCAGTTTGTGGTGAGTTGAGTTGCCATCCTTAGTTTTGTGGGTCTCACTCCG
MATCSCSPCLSSRCPSHS          233       ATGGCAACGTGTTCTTGTTCTCCGTGCCTGTCGTCTCGTTGTCCTTCTCACTCCG
MAICGTLPCASRRCPSHS          186       ATGGCAATTTGTGGTACGCTGCCGTGCGCGTCTCGTCGTTGTCCGTCTCACTCCG
MASCQLLSCGYSECSSHS          200       ATGGCATCTTGTTAGCTGCTTTCTTGCGGGTATTCGGAGTGTTCGTCTCACTCCG
MASCQVLGCWNSTCSSHS          104       ATGGCATCGTGTTAGGTTTTGGGGTGCTGGAATTCGACTTGTAGTTCTCACTCCG

 Group3
MAVCGVHLCSSGPCKSHS          373       ATGGCAGTTTGTGGGGTGCATTTGTGCTCTAGTGGTCCTTGTAAGTCTCACTCCG
MAICSLVVCRSGLCSSHS          140       ATGGCAATTTGTTCGCTTGTGGTGTGCCGGTCTGGGTTGTGTAGTTCTCACTCCG
MAYCRAGMCGQGPCSSHS          358       ATGGCATATTGTCGGGCGGGTATGTGCGGGTAGGGTCCGTGTTCGTCTCACTCCG
MACCQGGICGKGTCLSHS          206       ATGGCATGTTGTTAGGGTGGTATTTGCGGGAAGGGGACGTGTTTGTCTCACTCCG
MAQCVASICLYSNCSSHS          144       ATGGCACAGTGTGTTGCGTCTATTTGCTTGTATTCGAATTGTTCTTCTCACTCCG
MAYCAASICLQDICGSHS          138       ATGGCATATTGTGCTGCTTCTATTTGCTTGTAGGATATTTGTGGGTCTCACTCCG
MAECISSVCTRGVCLSHS          1791       ATGGCAGAGTGTATTTCTAGTGTGTGCACTAGGGGTGTTTGTTTGTCTCACTCCG
MASCFKDLCTRGLCNSHS          265       ATGGCATCGTGTTTTAAGGATCTGTGCACTAGGGGGCTGTGTAATTCTCACTCCG
MAVCMRVCVGNCGSHS            221       ATGGCAGTTTGTATGAGGGTTTGCGTGGGGAATTGTGGGTCTCACTCCG
MALCFRVCSGSCVSHS            148       ATGGCATTGTGTTTTCGTGTGTGCTCGGGTTCGTGTGTGTCTCACTCCG
MACCELVCGGSCDSHS            114       ATGGCATGTTGTGAGTTGGTTTGCGGGGGTTCTTGTGATTCTCACTCCG

 Group4
MARCRFTGCDIQVCLSHS          235       ATGGCACGGTGTAGGTTTACTGGTTGCGATATTTAGGTTTGTTTGTCTCACTCCG
MANCARLGCQVRLCLSHS          217       ATGGCAAATTGTGCTCGGTTGGGGTGCTAGGTGCGGTTGTGTTTGTCTCACTCCG
MASCACTGCFVPVCMSHS          185       ATGGCATCGTGTGCGTGTACTGGTTGCTTTGTGCCGGTGTGTATGTCTCACTCCG
MAGCRSSLCPWGMCFSHS          149       ATGGCAGGGTGTAGGTCTTCTTTGTGCCCTTGGGGTATGTGTTTTTCTCACTCCG

 Group5
MAVCVMRCGSTCGSHS          318       ATGGCAGTTTGTGTTATGCGGTGCGGGTCTACGTGTGGGTCTCACTCCG
MAICYVNCGRDCASHS          247       ATGGCAATTTGTTATGTGAATTGCGGTCGTGATTGTGCTTCTCACTCCG
MAICRLSCCHECGSHS          163       ATGGCAATTTGTCGTCTGTCTTGCTGTCATGAGTGTGGGTCTCACTCCG
MAVCLPKCESFCASHS          272       ATGGCAGTGTGTCTGCCTAAGTGCGAGAGTTTTTGTGCGTCTCACTCCG
MAICFEKCPNWCGSHS          193       ATGGCAATTTGTTTTGAGAAGTGCCCGAATTGGTGTGGTTCTCACTCCG
MALCRIPCESTCISHS          220       ATGGCACTTTGTAGGATTCCGTGCGAGAGTACTTGTATTTCTCACTCCG
MAGCRPPCDDTCGSHS          176       ATGGCAGGGTGTCGTCCTCCGTGCGATGATACGTGTGGTTCTCACTCCG
MAVCMGPCPGCCVSHS          562       ATGGCAGTGTGTATGGGTCCGTGCCCTGGGTGTTGTGTGTCTCACTCCG
MAQCTGKCGGCCSSHS          154       ATGGCATAGTGTACTGGGAAGTGCGGTGGGTGTTGTTCGTCTCACTCCG
MAHCSGVCPRCCPSHS          156       ATGGCACATTGTTCGGGGGTGTGCCCTAGGTGTTGTCCTTCTCACTCCG
MAVCLGDCPVDCASHS          255       ATGGCAGTGTGTCTTGGTGATTGCCCTGTGGATTGTGCGTCTCACTCCG
MATCLGSCGMSCTSHS          100       ATGGCAACTTGTCTTGGTTCTTGCGGGATGTCTTGTACTTCTCACTCCG

 Group6
MAHCILGCVPMCGSHS          607       ATGGCACATTGTATTTTGGGTTGCGTTCCGATGTGTGGGTCTCACTCCG
MAKCRLGCAPECGSHS          212       ATGGCAAAGTGTCGGCTGGGTTGCGCGCCTGAGTGTGGTTCTCACTCCG
MASCYGGCAGLCGSHS          355       ATGGCATCTTGTTATGGGGGGTGCGCTGGGCTTTGTGGGTCTCACTCCG
MAYCTLGCDTDCSSHS          718       ATGGCATATTGTACTTTGGGTTGCGATACGGATTGTTCTTCTCACTCCG
MAYCAYSCNSACTSHS          144       ATGGCATATTGTGCTTATTCGTGCAATTCGGCGTGTACGTCTCACTCCG
MADCAYGCDGECDSHS          332       ATGGCAGATTGTGCTTATGGGTGCGATGGGGAGTGTGATTCTCACTCCG
MATCRQGCVGACSSHS          718       ATGGCAACGTGTCGGTAGGGTTGCGTGGGTGCTTGTAGTTCTCACTCCG
MAGCRQGCSASCTSHS          161       ATGGCAGGTTGTAGGTAGGGGTGCAGTGCTTCTTGTACGTCTCACTCCG
MANCRQVGCGFARCSSHS        315       ATGGCAAATTGTCGGCAGGTGGGGTGCGGGTTTGCTCGTTGTAGTTCTCACTCCG
MANCRHGGCRGDHCTSHS        130       ATGGCAAATTGTAGGCATGGTGGTTGCCGGGGGGATCATTGTACGTCTCACTCCG

 Group7
MAVCGYDCSQMCSSHS          244       ATGGCAGTGTGTGGGTATGATTGCTCTTAGATGTGTTCTTCTCACTCCG
MAICGGACWQQCASHS          159       ATGGCAATTTGTGGGGGTGCTTGCTGGTAGCAGTGTGCCTCTCACTCCG
MAACGALCDPDCGSHS          274       ATGGCAGCTTGTGGTGCGTTGTGCGATCCGGATTGTGGGTCTCACTCCG
MAICGRLCTPLCRSHS          104       ATGGCAATTTGTGGTAGGCTTTGCACTCCTTTGTGTCGGTCTCACTCCG

 Group8
MASCVTVCLWTCGSHS          257       ATGGCATCTTGTGTTACTGTGTGCCTTTGGACTTGTGGGTCTCACTCCG
MACCNERCVWKCGSHS          241       ATGGCATGTTGTAATGAGCGGTGCGTGTGGAAGTGTGGGTCTCACTCCG
MAWCPGQCAWKCGSHS          104       ATGGCATGGTGTCCGGGGCAGTGCGCTTGGAAGTGTGGGTCTCACTCCG

 Group9
MAGCGRYCDGWCNSHS          325       ATGGCAGGGTGTGGGCGTTATTGCGATGGTTGGTGTAATTCTCACTCCG
MAVCGEHCDLWCYSHS          121       ATGGCAGTGTGTGGTGAGCATTGCGATTTGTGGTGTTATTCTCACTCCG
MALCAKSCSKWCRSHS          142       ATGGCACTGTGTGCTAAGAGTTGCTCGAAGTGGTGTAGGTCTCACTCCG

 Group10
MAVCRSRCVMYCQSHS          428       ATGGCAGTGTGTCGGTCTCGGTGCGTGATGTATTGTTAGTCTCACTCCG
```

```
MAVCATQCMTMCQSHS          266      ATGGCAGTTTGTGCTACTTAGTGCATGACGATGTGTTAGTCTCACTCCG
MAVCVQSCFIKCQSHS          212      ATGGCAGTTTGTGTGTAGAGTTGCTTTATTAAGTGTCAGTCTCACTCCG

 Group11
MAECMDCCSEECLSHS          106      ATGGCAGAGTGTATGGATTGTTGCTCTGAGGAGTGTTTGTCTCACTCCG
MASCFDCCGWLCNSHS          305      ATGGCATCGTGTTTTGATTGTTGCGGGTGGCTGTGTAATTCTCACTCCG
MASCELCCDVHCYSHS          178      ATGGCATCTTGTGAGTTGTGTTGCGATGTTCATTGTTATTCTCACTCCG
MAFCQVCCHPLCYSHS          407      ATGGCATTTTGTTAGGTTTGTTGCCATCCGTTGTGTTATTCTCACTCCG
MALCAVKCHWYCRSHS          202      ATGGCATTGTGTGCGGTTAAGTGCCATTGGTATTGTAGGTCTCACTCCG
MAICAMVCGWFCASHS          111      ATGGCAATTTGTGCTATGGTGTGCGGTTGGTTTTGTGCTTCTCACTCCG
MAKCAVWCQLYCPSHS          460      ATGGCAAAGTGTGCGGTGTGGTGCTAGTTGTATTGTCCTTCTCACTCCG
MAVCSWECRVFCASHS          204      ATGGCAGTGTGTAGTTGGGAGTGCAGGGTGTTTTGTGCGTCTCACTCCG
MAGCSAWECRVLCLSHS         167      ATGGCAGGGTGTAGTGCGTGGGAGTGCAGGGTTCTGTGTTTGTCTCACTCCG
MAGCSWECMAWCWSHS          382      ATGGCAGGTTGTTCGTGGGAGTGCATGGCTTGGTGTTGGTCTCACTCCG
MAGCKWVCDSLCVSHS          112      ATGGCAGGGTGTAAGTGGGTGTGCGATTCTTTGTGTGTTTCTCACTCCG
MAGCALVCTRFCLSHS          132      ATGGCAGGTTGTGCGCTGGTGTGCACGCGTTTTTGTCTGTCTCACTCCG
MASCPGMCRGSCTSHS          102      ATGGCAAGTTGTCCGGGGATGTGCCGTGGGAGTTGTACTTCTCACTCCG
MASCSGFCLMPCNSHS          354      ATGGCAAGTTGTTCGGGGTTTTGCCTTATGCCGTGTAATTCTCACTCCG
MAACSKRCLHPCLSHS          325      ATGGCAGCGTGTTCTAAGCGGTGCCTTCATCCTTGTCTTTCTCACTCCG
MAACSKLRCVMHVCESHS        217      ATGGCAGCGTGTAGTAAGTTGAGGTGCGTTATGCATGTTTGTGAGTCTCACTCCG

 Group12
MAGCQELWCQVAHCWSHS          390      ATGGCAGGGTGTTAGGAGCTTTGGTGCCAGGTGGCTCATTGTTGGTCTCACTCCG
MATCLWAVCLMSVCLSHS          113      ATGGCAACTTGTCTGTGGGCGGTGTGCTTGATGTCGGTTTGTCTTTCTCACTCCG
MADCQWRCCLSRGCLSHS          201      ATGGCAGATTGTCAGTGGCGTTGTTGCTTGTCGAGGGGTTGTCTGTCTCACTCCG
MAWCAWFSCATTDCGSHS          161      ATGGCATGGTGTGCGTGGTTTTCGTGCGCGACGACGGATTGTGGGTCTCACTCCG
MAWCQWWYCQTHECLSHS          125      ATGGCATGGTGTCAGTGGTGGTATTGCTAGACTCATGAGTGTTTGTCTCACTCCG

 Group13
MARCWGEGCRGLLCHSHS          119      ATGGCAAGGTGTTGGGGTGAGGGTTGCAGGGGTTTGTTGTGTCATTCTCACTCCG
MADCWDFACSGVVCPSHS          166      ATGGCAGATTGTTGGGATTTTGCTTGCTCTGGTGTTGTGTGTCCTTCTCACTCCG
MAGCTGFGCSGQICSSHS          326      ATGGCAGGTTGTACGGGGTTTGGGTGCTCGGGTTAGATTTGTTCGTCTCACTCCG
MAQCTAVGCHGGVCGSHS          568      ATGGCATAGTGTACGGCGGTGGGGTGCCATGGTGGTGTGTGTGGTTCTCACTCCG
MASCLAYTCRARVCPSHS          164      ATGGCAAGTTGTTTGGCTTATACGTGCAGGGCGCGTGTTTGTCCTTCTCACTCCG
MAPCLLQWCRLRVCKSHS          170      ATGGCACCTTGTTTGCTGTAGTGGTGCCGTTTGCGTGTGTGTAAGTCTCACTCCG

 Group14
MARCQTKSCLKWVCSSHS          123      ATGGCACGTTGTTAGACTAAGAGTTGCCTTAAGTGGGTTTGTAGTTCTCACTCCG
MAGCQTRLCLGWECTSHS          192      ATGGCAGGGTGTTAGACTCGTTTGTGCCTGGGTTGGGAGTGTACTTCTCACTCCG
MASCHPQVCLLWECGSHS          112      ATGGCATCGTGTCATCCTTAGGTGTGCCTTTTGTGGGAGTGTGGTTCTCACTCCG
MACCQTRVCLVEFCGSHS          191      ATGGCATGTTGTTAGACTCGGGTGTGCCTTGTGGAGTTTTGTGGGTCTCACTCCG
MANCSTVTCMELSCTSHS          102      ATGGCAAATTGTAGTACTGTTACGTGCATGGAGCTTTCTTGTACTTCTCACTCCG
MANCQTMHCLRGDCSSHS          456      ATGGCAAATTGTTAGACTATGCATTGCTTGCGTGGTGATTGTAGTTCTCACTCCG
MAGCPGKICWSGVCGSHS          302      ATGGCAGGGTGTCCTGGTAAGATTTGCTGGTCGGGTGTTTGTGGGTCTCACTCCG
MAGCPEAACCLQVCGSHS          754      ATGGCAGGGTGTCCGGAGGCGGCTTGCTGTCTGTAGGTTTGTGGGTCTCACTCCG
MALCYSALCSGVGCGSHS          100      ATGGCACTTTGTTATTCTGCTCTGTGCTCGGGTGTGGGGTGTGGGTCTCACTCCG
MALCCHAVCDDGVCGSHS          191      ATGGCACTGTGTTGTCATGCGGTTTGCGATGATGGTGTTTGTGGGTCTCACTCCG

 Group15
MALCQGVDCVFFICFSHS          265      ATGGCATTGTGTTAGGGTGTTGATTGCGTTTTTTTTATTTGTTTTTCTCACTCCG
MAQCAGTACMFWTCESHS          490      ATGGCATAGTGTGCTGGGACGGCGTGCATGTTTTGGACTTGTGAGTCTCACTCCG
MAECVGGMCHMIACCSHS          412      ATGGCAGAGTGTGTGGGGGGTATGTGCCATATGATTGCTTGTTGTTCTCACTCCG
MAVCEGCFCQMLICCSHS          274      ATGGCAGTGTGTGAGGGTTGTTTTTGCTAGATGTTGATTTGTTGTTCTCACTCCG

 Group16
MAGCCFPTCSWLVCLSHS          474      ATGGCAGGTTGTTGTTTTCCTACTTGCTCGTGGCTGGTGTGTCTTTCTCACTCCG
MAQCCLSQCDWLVCLSHS          167      ATGGCATAGTGTTGTCTGAGTCAGTGCGATTGGTTGGTTTGTCTTTCTCACTCCG
MASCKAGACVQKWCVSHS          223      ATGGCAAGTTGTAAGGCTGGTGCGTGCGTTTAGAAGTGGTGTGTTTCTCACTCCG
MARCVMGLCEQVLCISHS          100      ATGGCACGTTGTGTTATGGGGTTGTGCGAGTAGGTGTTGTGTATTTCTCACTCCG
MATCRDSSCEQLICQSHS          454      ATGGCAACGTGTCGTGATAGTAGTTGCGAGCAGCTTATTTGTTAGTCTCACTCCG
MAWCKAGLCHGNFCLSHS          134      ATGGCATGGTGTAAGGCGGGGTTGTGCCATGGTAATTTTTGTTTGTCTCACTCCG
MAWCYQGKCVGLLCASHS          220      ATGGCATGGTGTTATTAGGGTAAGTGCGTTGGTCTGCTTTGTGCGTCTCACTCCG

 Group17
MAVCHPQFCSGVLCYSHS          113      ATGGCAGTGTGTCATCCGTAGTTTGCTCGGGGGTTTTGTGTTATTCTCACTCCG
MAPCHPQVCRRELCVSHS          132      ATGGCACCGTGTCATCCTCAGGTGTGCAGGAGGGAGCTTTGTGTTTCTCACTCCG
MAECHPQNCPSSWCLSHS          101      ATGGCAGAGTGTCATCCTTAGAATTGCCCTAGTTCTTGGTGTTTGTCTCACTCCG
MAFCHPQNCSGSDCISHS          198      ATGGCATTTTGTCATCCGCAGAATTGCAGTGGTTCTGATTGTATTTCTCACTCCG
MAGCHPQNCVLASCMSHS          124      ATGGCAGGGTGTCATCCGTAGAATTGCGTTCTTGCGAGTTGTATGTCTCACTCCG
MASCCPQKCTESMCRSHS          207      ATGGCAAGTTGTTGTCCTTAGAAGTGCACGGAGAGTATGTGTCGGTCTCACTCCG

 Group MIXED
MATCPVPLCRVMPCRSHS          259      ATGGCAACTTGTCCTGTGCCTTTGTGCCGGGTGATGCCGTGTAGGTCTCACTCCG
MALCPSTQCRTVPCVSHS          206      ATGGCATTGTGTCCGAGTACGTAGTGCCGGACTGTTCCGTGTGTTTCTCACTCCG
MARCFSWECWVFGCSSHS          195      ATGGCACGTTGTTTTAGTTGGGAGTGCTGGGTGTTTGGGTGTAGTTCTCACTCCG
MAVCLSWECWVFGCQSHS          147      ATGGCAGTTTGTTTGAGTTGGGAGTGCTGGGTTTTTGGGTGTTAGTCTCACTCCG
MAWCVLIDWFCSHS              303      ATGGCATGGTGTGTACTAATTGACTGGTTTTGTTCTCACTCCG
MADCGHPQFRVYVCPSHS          242      ATGGCAGATTGTGGGCATCCTTAGTTCCGGGTTTATGTTTGTCCGTCTCACTCCG
```

MAACTDNCFEWCFSHS	251	ATGGCAGCGTGTACGGATAATTGCTTTGAGTGGTGTTTTTCTCACTCCG
MAACTPDNCMEPHCTSHS	255	ATGGCAGCGTGTACGCCTGATAATTGCATGGAGCCGCATTGTACCTCTCACTCCG
MAVCPRWCMDPCKSHS	132	ATGGCAGTTTGTCCGCGGTGGTGCATGGATCCGTGTAAGTCTCACTCCG
MAECCQGCIILCDSHS	311	ATGGCAGAGTGTTGTTAGGGGTGCATTATTTTGTGTGATTCTCACTCCG
MADCGSVCPHMCFSHS	103	ATGGCAGATTGTGGGTCGGTTTGCCCGCATATGTGTTTTTCTCACTCCG
MAGCGWGWCFTWPCISHS	199	ATGGCAGGTTGTGGGTGGGGTTGGTGCTTTACTTGGCCTTGTATTTCTCACTCCG
MAGCGAWGCSVYISMSHS	133	ATGGCAGGGTGTGGTGCGTGGGGTTGCAGTGTTTATATTTCTATGTCTCACTCCG
MAWCWGLWCSYGWCDSHS	218	ATGGCATGGTGTTGGGGGCTTTGGTGCAGTTATGGGTGGTGTGATTCTCACTCCG
MAACQQIWCNFNSCMSHS	160	ATGGCAGCTTGTCAGTAGATTTGGTGCAATTTTAATTCTTGTATGTCTCACTCCG
MAVCSPASCWFKCCQSHS	101	ATGGCAGTGTGTTCGCCGGCGTCTTGCTGGTTTAAGTGTTGTTAGTCTCACTCCG
MAGCHQGVCSPSLCESHS	369	ATGGCAGGGTGTCATTAGGGTGTGTGCAGTCCTTCTTTGTGTGAGTCTCACTCCG
MAGCMRGVCPVGDCGSHS	161	ATGGCAGGGTGTATGCGGGGGGTGTGCCCTGTGGGTGATTGTGGTTCTCACTCCG
MARCPCSVCYDMHCGSHS	227	ATGGCACGGTGTCCGTGTAGTGTGTGCTATGATATGCATTGTGGTTCTCACTCCG
MANCNGRPCALGGCGSHS	104	ATGGCAAATTGTAATGGGCGGCCGTGCGCGTTGGGTGGGTGTGGGTCTCACTCCG
MAECRQEPCAQWGCYSHS	103	ATGGCAGAGTGTAGGTAGGAGCCGTGCGCTCAGTGGGGTTGTTATTCTCACTCCG
MAYCRILDCGMSKCRSHS	128	ATGGCATATTGTCGTATTCTGGATTGCGGGATGAGTAAGTGTCGGTCTCACTCCG
MARCRMEVCPYGRCQSHS	316	ATGGCAAGGTGTCGTATGGAGGTTTGCCCGTATGGTAGGTGTCAGTCTCACTCCG
MAWCGRQACVVGECKSHS	176	ATGGCATGGTGTGGGCGGTAGGCGTGCGTTGTGGGGGAGTGTAAGTCTCACTCCG
MAYCHLQACQDGCCLSHS	103	ATGGCATATTGTCATTTGCAGGCGTGCTAGGATGGTTGTTGTTTGTCTCACTCCG
MASCGSVPCGVMHCVSHS	590	ATGGCATCGTGTGGGTCTGTTCCTTGCGGGGTTATGCATTGTGTGTCTCACTCCG
MAWCVQMGCGDLHCVSHS	369	ATGGCATGGTGTGTTTAGATGGGGTGCGGTGATTTGCATTGTGTTTCTCACTCCG
MAVCANRFCSVAACDSHS	196	ATGGCAGTTTGTGCGAATCGGTTTTGCAGTGTTGCGGCTTGTGATTCTCACTCCG
MAFCDGVFCFRAPCFSHS	127	ATGGCATTTTGTGATGGGGTTTTTTGCTTTAGGGCTCCGTGTTTTTCTCACTCCG
MALCGENGCPLSVCVSHS	107	ATGGCATTGTGTGGTGAGAATGGGTGCCCGTTGAGTGTGTGTGTGTCTCACTCCG
MALCLENCCNMFECRSHS	154	ATGGCATTGTGTCTTGAGAATTGTTGCAATATGTTTGAGTGTCGGTCTCACTCCG
MAHCYRKHCDMTDCPSHS	200	ATGGCACATTGTTATCGGAAGCATTGCGATATGACTGATTGTCCGTCTCACTCCG
MAKCASLGCVLSSCPSHS	239	ATGGCAAAGTGTGCGTCGCTGGGTTGCGTTTTGTCTAGTTGTCCTTCTCACTCCG
MAHCAYLRCTLLLCQSHS	798	ATGGCACATTGTGCGTATTTGCGGTGCACGTTGTTGTTGTGTTAGTCTCACTCCG
MARCLERLCGCRVCFSHS	244	ATGGCACGGTGTCTGGAGCGGCTGTGCGGGTGTCGTGTGTGTTTTTCTCACTCCG
MACCVGEPCSPVVCPSHS	488	ATGGCATGTTGTGTGGGTGAGCCGTGCTCGCCGGTTGTTTGTCCTTCTCACTCCG
MAVCFTGTCPWWFCFSHS	294	ATGGCAGTGTGTTTTACGGGTACGTGCCCGTGGTGGTTTTGTTTTTCTCACTCCG
MASCGSGECQVWFCPSHS	284	ATGGCATCGTGTGGGTCGGGGGAGTGCTAGGTTTGGTTTTGTCCGTCTCACTCCG
MAECDLSDCWIVNCFSHS	106	ATGGCAGAGTGTGATCTTAGTGATTGCTGGATTGTGAATTGTTTTTCTCACTCCG
MAECWQLECYWLMCGSHS	647	ATGGCAGAGTGTTGGTAGTTGGAGTGCTATTGGTTGATGTGTGGGTCTCACTCCG
MAMCWFSSCWLGICLSHS	137	ATGGCAATGTGTTGGTTTAGTAGTTGCTGGTTGGGTATTTGTTTGTCTCACTCCG
MAKCWRGDCALSVCASHS	1064	ATGGCAAAGTGTTGGCGTGGGGATTGCGCTCTGTCTGTGTGTGCGTCTCACTCCG
MAHCWRKTCTGDMCLSHS	138	ATGGCACATTGTTGGCGTAAGACGTGCACGGGGGATATGTGTTTGTCTCACTCCG
MACRGCGSHS	223	ATGGCATGTCGTGGGTGTGGTTCTCACTCCG
MALYIPSILSHS	184	ATGGCTTTGTATATTCCTTCTATTTTGTCTCACTCCG

MatLab Scripts description

Step1.m

- It reads the initial filename.fastq file and generates files containing reads according to their barcode (named BC1.txt, BC2.txt... BCNOT.txt), and saves them in a separate folder within the input folder (named "filename_BC"). BCNOT.txt contains reads whose barcode did not correspond to any of the identified barcodes. If no input is indicated, a dialog box opens to choose the file and the barcodes used are the 12 described in this publication.
- Input: (optional)
    - Step1('inname','filename.fastq','indir','path') indicates the file name and path to the folder where it is located. If not specified, a dialog box to choose the file will open.
    - Step1(...,'indelmut','on') allows one insertion, deletion or mutation in the barcodes. If not specified, it is off.
    - Step1(...,'bc',{'AAAAAA',' TTTTTT','GGGGGG';'...'}) indicates the barcodes used. They must be separated by comma, in single braket and within {}. If not specified, it uses the ones described in Table S2.
- Output:
    - Command window:
        - chip-specific code
        - time taken to read different fractions of filename.fastq
    - A new folder called "filename_BC" with a series of files named BC1.txt, BC2.txt... BCn.txt, containing the reads corresponding to the first, second,... nth barcode respectively. Reads whose barcode did not correspond to any of the identified barcodes are stored in BCNOT.txt.
    - BC_stats.txt file, containing information about how many reads where found per barcode.

Step2.m

- It removes low quality reads from the datasets, groups identical DNA sequences and sorts them by abundance. It then translates them (amber codon is translated to glutamine). If no input is specified, a dialog box opens that allows choosing the file (a BCn.txt output of Step1). The constant DNA sequences flanking the random region must be indicated (if not indicated, start and end of the random region are the ones suitable for bicyclic peptide libraries used in this publication). Default quality parameters are 3 base calls below Q18. Optionally, a minimum, intermediate and maximum length of the peptide can be indicated. The additional correcting error step is recommended for low-diversity datasets in which a few clones predominate in the library, but it may take 10-20 minutes. It will merge together sequencing having one or two different positions in the DNA sequence.
- input (optional):
    - Step2('inname','filename.txt','indir','path') indicates the file name and path to the folder where it is located. If not specified, a dialog box to choose the file will open. Step2 can read the output file from Step1 (BC1.txt, BC2.txt...).
    - Step2(...,'badmax',n) where n is the maximum number of bases below the quality threshold allowed. If not specified, badmax = 3
    - Step2(...,'q',Q) where Q indicates the quality threshold (18 for Q18, 20 for Q20, etc...). If not specified, Q = 18.
    - Step2(...,'uplimit',m,'downlimit',o,'midlimit',p) specifies the maximum (m) and minimum (o) peptide length (in residues). Additionally, an internal limit can be indicated (p).
    - Step2(...,'start','NNKNNK','end','NNKNNK') specifies constant regions at the start and end of the DNA region of interest. If not specified, it uses the ones described in this publication. The first nucleotide of 'start' must be the first nucleotide of the codon for the translation to be in frame.
    - Step2(...,'fixerr',n) allows to correct sequencing errors: it merges together sequencing with only 1 or 2 differences in the DNA sequence. It corrects only the top "n" abundant sequences.
- output:
    - Translation_filename folder containing the file Translated_filename_GOOD.txt and Translated_stats.txt (indicating the number of different sequences, maximum abundance and total number of reads).
    - Optionally: additional files folder within the previous folder containing the translation files of the bad quality reads, too long reads and too short reads.
    - Optionally: QF_filename folder with QF_filename_GOOD.txt and QF_filename_BAD.txt containing good and bad quality reads respectively. QF_filename_NOLIM.txt contains reads where either the start or the end of the region of interest could not be found. QF_filename_toolong.txt and

QF_filename_tooshort.txt contain the reads whose peptides where shorter or longer than the limits indicated. If an intermediate limit was indicated, two files: a QF_filename_longGOOD.txt and a QF_filename_shortGOOD.txt, are created. IDEM with BAD.

- If 'fixerr' option is on, within the Translation_filename folder, a file called "fixerrTranslated_filename.txt" appears, as well as an additional folder "correction data". In this folder there are files with all the correction events (potential conflict ones are in a separate file for an easier evaluation, i.e. if the abundances differ in less than 4-fold). ErrorRates.txt contains the error rate (1st column), the starting occurrence (2nd column) and the final occurrence after correcting (3rd column) of the peptides.

LoopLengths.m (for monocyclic and bicyclic peptide libraries)

- Separates sequences in different files according to the peptide format (i.e. number of cysteines and the number of residues between them). If no input is specified, a dialog box opens that allows choosing the file (having the format: peptide seq - abundance - nucleotide seq). Optionally, one can indicate a minimum abundance for a sequence to be considered, and a constant C-terminal peptide sequence to remove frame-shifted sequences.
- Input (optional):
    - LoopLengths('inname','filename.txt','indir','path') indicates the file name and path to the folder where it is located. If not specified, a dialog box to choose the file will open. LoopLengths needs a file with data on the format: peptide seq - abundance - nucleotide seq.
    - LoopLengths(...,'cutoff',n), where n specifies the minimum abundance to be considered. If not specified, no cutoff is applied and all sequences are considered.
    - LoopLengths(...,'cter','XXX'), where XXX is the amino acid sequence found at the C-terminus of the peptide. Allows to remove frame-shifted clones.
- Output:
    - A new folder called "LoopLengths_filename" with 4 subfolders:
        - 2cys: containing files with the sequences corresponding to 2 cysteines, subdivided by loop lenght. Example of the notation:
          3_twocys = C XXX C
          5_twocys = C XXXXX
        - 3cys: containing files with the sequences corresponding to 3 cysteines, subdivided by loop lenght. Example of the notation:
          3_threecys = 0x3 = CC XXX C
          300_threecys = 3x0 = C XXX CC
          304_threecys = 3x4 = C XXX C XXXX C
          305_threecys = 3x5 = C XXX C XXXXX C
        - 4cys: containing files with the sequences corresponding to 4 cysteines, subdivided by loop lenght. Example of the notation:
          4_fourcys = 0x0x4 = CCC XXXX C
          400_fourcys = 0x4x0 = CC XXXX CC
          403_fourcys = 0x4x3 = CC XXXX C XXX C
          40000_fourcys = 4x0x0 = C XXXX CCC
          40302_fourcys = 4x3x2 = C XXXX C XXX C XX C$
        - other: containing files with the sequences corresponding to 0, 1 or more than 4 cysteines; and "stats" file with the information about the number of total and different sequences assigned to each category.

Clustering.m

- Compares a chosen number of sequences (if not specified, compares top 200) and groups them into families that share high sequence similarity. Within a cluster, more similar sequences appear together. A figure logo for each group is generated and saved as a .jpg file within the input folder. Optionally, the number of different sequences OR the minimum abundance can be indicated. If no input is indicated, a dialog box will open that allows choosing the file. It must be a file of the format: format peptide seq. – abundance - nucleotide seq. Additionally, two optional parameters allow fine-tuning of the clustering: "min_clustersize" and "stringency".
- Input (optional):
    - Clustering('inname','filename.txt','indir','path') indicates the file name and path to the folder where it is located. If not specified, a dialog box to choose the file will open. Clustering can read the output files from Step2, LoopLengths and FindSeq. Requisites: data is on the format peptide seq – abundance - nucleotide seq.
    - Clustering(...,'number_dif',n) indicates how many different sequences will be clustered. If not specified, n = 200, i.e. top 200 most abundant sequences will be clustered.
    - Clustering(...,'min_clustersize',m) indicates the minimum number of sequences within a cluster to be considered. If a cluster has less than m sequences, it will be transferred to the "mixed" cluster.
    - Clustering(...'cter','XXXX') indicates a constant C-terminal region of the peptide. Peptides without this constant region will be not considered.
    - Clustering(...,'min_abun',n) indicates the minimum abundance for clones to be considered.
    - Clustering(...,'stringency',s) allows to fine-tune the clustering of the script to different datasets. In general, higher values of stringency will lead to more similar peptides within each cluster and more sequences in the mixed cluster. Lower values of stringency allow more differences within each cluster and as a result fewer sequences go to the mixed cluster.
    - Clustering(...,'logos','off') disables the generation of sequence logos (.jpg files) within the input folder.
    - Clustering(...'gappen',n) changes the value of gap opening and gap extension penalties. Default value is 8.
- Output:
    - Clusters_filename.txt file within the same folder as the input file.
    - A series of .jpg files corresponding to the sequence logos of each group within the same folder as the input file.

FindSeq.m

- Searches the dataset for all peptide sequences containing a specified motif. The motif must be specified in the input, and can be a string of characters or a regular expression. It distributed the peptides in two different files, according to whether they contain the specified motif or not.
- Input:
    - FindSeq('seq','XXX'), will look for XXX motif. For example, FindSeq('seq','HPQ') will look for all sequences in the dataset containing HPQ. Regular expressions can be used instead, for example, FindSeq('seq','H.Q') will look for all sequences conatining HXQ, X being any amino acid. FindSeq('seq','H.?Q') will look for all sequences containing HQ or HXQ, X being any amino acid. For more informaiton about regular expressions, see MatLab help.
- Input (optional):
    - FindSeq('inname','filename.txt','indir','path') indicates the file name and path to the folder where it is located. If not specified, a dialog box to choose the file will open. Requisites: data is on the format peptide seq – abundance - nucleotide seq.
    - FindSeq(...'cter','XXXX') indicates a constant C-terminal region of the peptide. Peptides without this constant region will be not considered.
    - FindSeq(...,'cutoff',n) indicates the minimum abundance for clones to be considered.
- Output:
    - A new folder named "Seq" within the input folder, containing three files: Seq_XXX_match.txt (XXX is the specified regular expression where especial characters have been substituted by "_"), containing all sequences that match the expression. Seq_XXX_nomatch.txt, containing the sequences that do not match it. And Seq_XXX_stats.txt, containing how many total and different sequences were assigned to each file.

CommonSeq.m

- Compares up to three different datasets and distributes common and exclusive sequences in different files.
- Input (optional):
    - CommonSeq('inname1','filename1.txt','inname2','filename2.txt','inname 3','filename3.txt','indir1,'path1','indir2','path2','indir3','path3') specifies three files and three paths corresponding to them. If not specified, dialog boxes will open for each.
    - CommonSeq(...,'cutoff',n), where n is the minimum abundance to be considered
    - CommonSeq(...,'top',m), alternative to the previous one, it indicates the top m abundant sequences of each file will be considered
    - CommonSeq(...,'cter','XXX'), specifies constant C-terminal residues (allows the removal of frame-shifted clones that do not have them)
- Output
    - A new folder named "comparison" within the folder containing the FIRST FILE. The following files are generated:
        - Comparison_seq1.txt, Comparison_seq2.txt, Comparison_seq3.txt = contain sequences that appeared only in the first, second and third file respectively
        - Comparison_seq12.txt, Comparison_seq13.txt, Comparison_seq23.txt = contain sequences that appeared in two of the files
        - Comparison_seq123.txt = contains sequences that appeared in the three files
        - Comparison_stats.txt = contains the number of total/different sequences considered in each case and the number of different sequences assigned to each file

APPENDIX III. Supplementary Information for Chapter 4

Supplementary experimental procedures

SrtA production

The vector pHTT14 for recombinant expression of $(His)_6$-$SrtA_{26\text{-}206}$ in *E. coli* was kindly provided by Prof. Olaf Schneewind (University of Chicago, IL, US). SrtA was produced as previously described. Briefly, pHTT14 was transformed in *E. coli* Xl1-blue, and was grown in LB (100 µg/ml ampicillin) until OD_{600} = 0.5. SrtA expression was induced by addition of 1 mM IPTG, and protein production was allowed for 4 h at 25 °C. Cells were harvested by centrifugation and resuspended in cold lysis buffer (30 mM NaH_2PO_4, 300 mM NaCl, 0.1% Triton X-100, 50 µg/ml DNAse, 100 µg/ml lysozyme, pH 7.4), and sonicated. Lysates were cleared by centrifugation and $(His)_6$-$SrtA_{26\text{-}206}$ was purified by Nickel-column chromatography followed by size exclusion chromatography (Figure S13).

Ion Torrent sequencing

Phage vector was isolated from TG1 *E. coli* after infection with a commercial plasmid purification kit (NucleoSpin Plasmid; Macherey-Nagel, Düren, Germany), and samples were prepared as previously described (Chapter 3), 100 ng phage vector DNA was amplified by PCR using primers containing adapter sequences and barcodes. The concentration of DNA was determined using a High Sensitivity DNA Assay Kit (Agilent, Santa Clara, CA, US), following the manufacturer's protocol. Ion Torrent sequencing was performed by the Lausanne Center of Genomic Technologies (University of Lausanne, Switzerland) on an Ion Personal Genome Machine (PGM™) Sequencer, using an Ion Torrent 316™ chip.

Determination of inhibitory constants (K_i)

For IC_{50}s higher than 10 µM, the inhibitory constant K_i was calculated according to the equation of Cheng and Prusoff[140] $K_i = IC_{50}/(1 + [S]_0/Km)$, wherein IC_{50} is the functional strength of the inhibitor, $[S]_0$ is the total substrate concentration, and Km is the Michaelis–Menten constant. The reported Km values for LPETG substrates of SrtA are between 5 and 7 mM, and therefore $K_i \approx IC_{50}$. For IC_{50}s lower than 10 µM, a second series was performed using 1 µM enzyme and 50 µM substrate, and K_is were calculated accordingly. For IC_{50}s lower than 3 µM in this second series, data was fitted to the Morrison equation:

$$V_i/V_0 = 1 - \frac{E_0 + I_0 + K_i - \sqrt{(E_0 + I_0 + K_i)^2 - 4E_0I_0}}{2E_0}$$

, where V_i and V_0 are the reaction velocities in the presence and absence of inhibitor, respectively. E_0 and I_0 represent the total enzyme and inhibitor concentration, respectively. K_i is the inhibition constant in the presence of fluorogenic substrate. Fitting curves were generated using OriginPro 8G software (OriginLab Corporation).

Determination of dissociation constants (K_d)

The dissociation constants (K_d) were determined by non-linear regression analyses of fluorescence polarization (F_p) versus total concentration of SrtA (P_T) using the following equation:

$$F_p = F_{p\,min} + (F_{p\,max} - F_{p\,min}) \frac{L_T + P_T + K_D - \sqrt{(L_T + P_T + K_D)^2 - 4L_TP_T}}{2L_T}$$

$F_{p\,min}$ and $F_{p\,max}$ are the fluorescence polarization for the free peptide and the fully bound peptide respectively, and L_T is the total concentration of fluorescent ligand (200 nM). Fitting curves were generated using OriginPro 8G software (OriginLab Corporation).

Supplementary figures

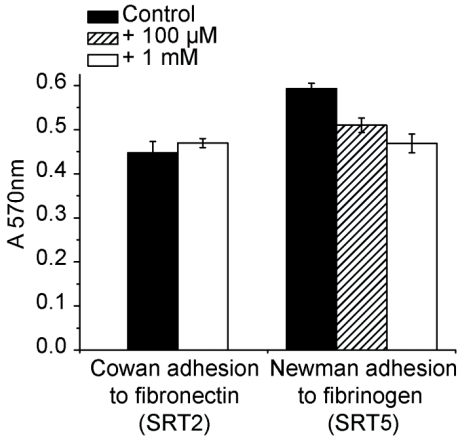


Figure S9. Adhesion tests using different bicyclic peptides (SRT2: ACTQRCPQLPPCG. SRT5: ACPLLPPCADDCG). Cultures of *S. aureus* strains Cowan and Newman were grown until mid-exponential phase (OD_{600} = 0.5) in the presence or absence of inhibitor. Adhesion test was performed immobilizing 1 µg of ligand protein per well (fibronectin or fibrinogen) and 5×10^7 cells/well. Average and standard deviations of two independent cultures are shown.

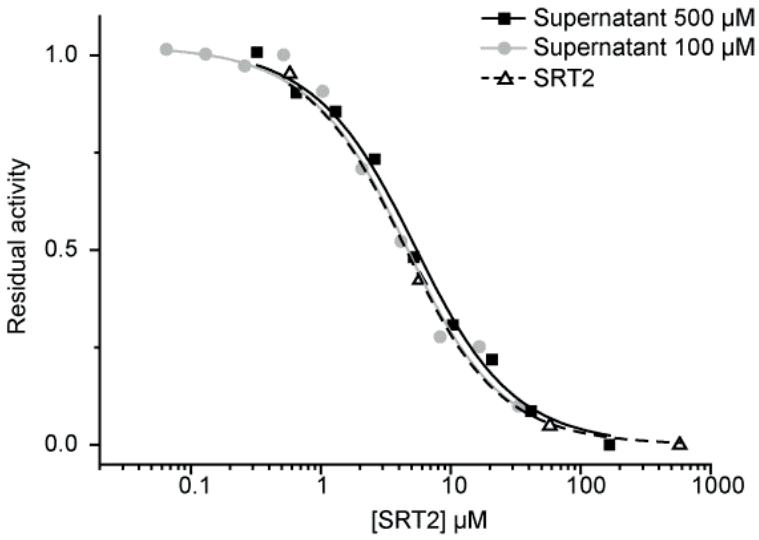


Figure S10. SRT2 is not degraded by proteases during *S. aureus* growth. Overnight supernatants of *S. aureus* Newman cultures containing 100 or 500 µM SRT2 were used for an activity assay in vitro. Inhibitory activity corresponded well to the theoretically expected if all peptide was intact (plotted x values in the graph). The presence of 40% supernatant in the in vitro activity assay did not affect SRT2 inhibitory activity (dashed line, K_i = 3.1 +/- 0.6 µM).

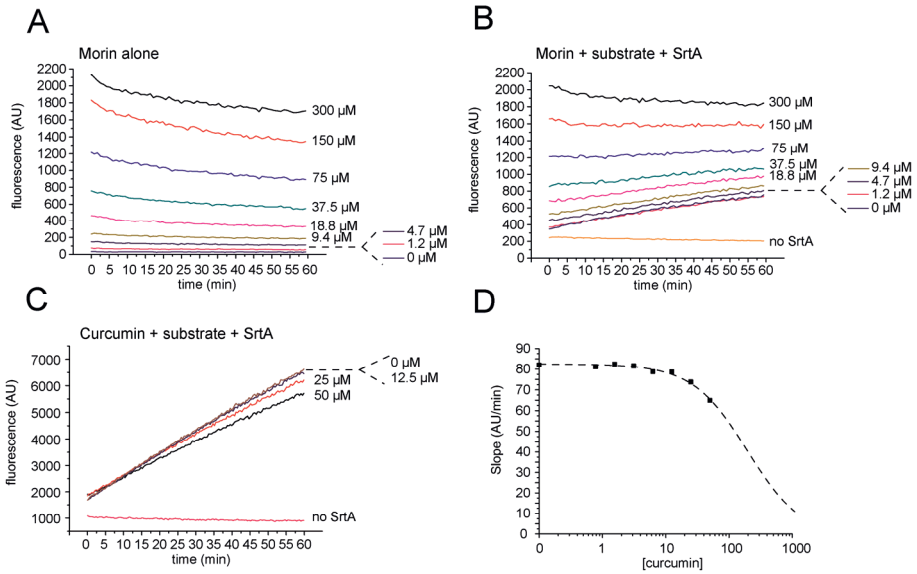


Figure S11. (A) Change of morin's fluorescence (ex. 350nm, em. 480nm) over time. (B) and (C) SrtA activity assays in the prescence of morin (B) or curcumin (C) using 2.5 µM SrtA and 20 µM Dab-LPETG-edans substrate, fluorescence (ex. 350nm, em. 480nm) was monitored for 1 h at 37 °C. (D) The IC_{50} of curcumin in our assays could not be precisely determined due to insolubility of the compound. At the maximum concentration tested (50 µM), 20% inhibition was observed.

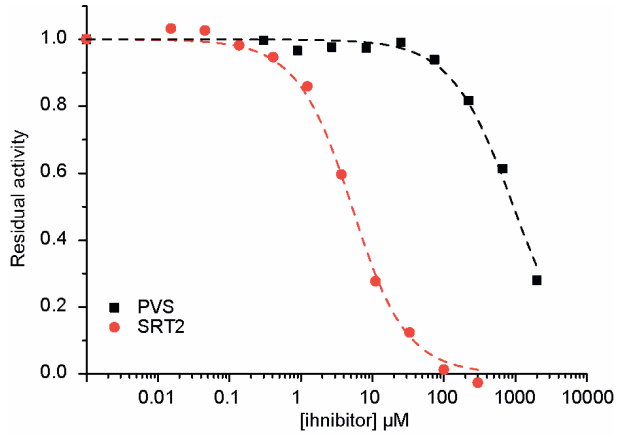


Figure S12. Inhibition of SrtA by phenyl-vinyl-sulfone (PVS). PVS was incubated at different concentrations with 2.5 µM SrtA for 1 h (1.5% DMSO final concentration) prior to the addition of Dabcyl-LPETG-Edans substrate. Cleavage was monitored for 1 h at 37 °C (IC_{50} = 0.9 mM). For comparison, a parallel assay was performed with SRT2.

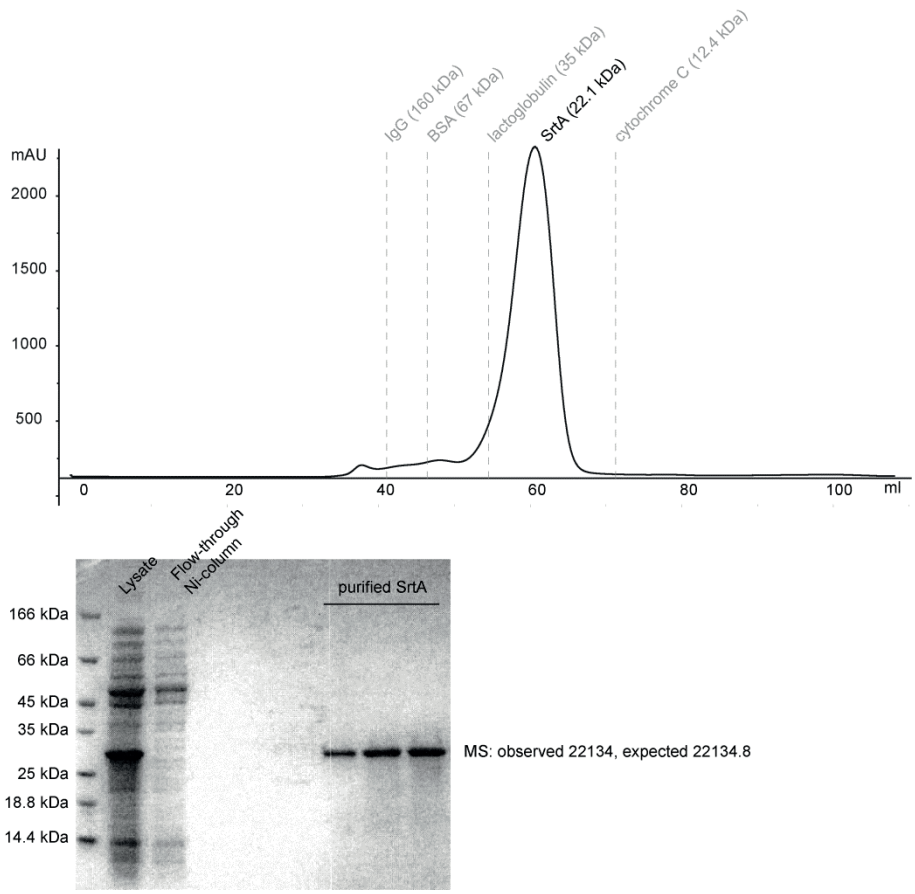


Figure S13. Purification of His6-SrtA$_{\Delta N}$. Top: size exclusion chromatography after nickel-column purification. Bottom: SrtA runs in the SDS-PAGE gel between 25 and 35 kDa. The mass of the purified fractions after size exclusion chromatography was verified by MS, and corresponded to the expected size.

Sequencing results (most abundant 100 clones)

Peptide sequence, abundance and nucleotide sequence are indicated.

Library A

MAACRQLPPCSFECGGSG	5059	ATGGCAGCATGCAGGTAGCTTCCTCCTTGCTCTTTCGAGTGTGGCGGTTCTGGCG
MAACQLLPPCPFNCGGSG	1514	ATGGCAGCATGCTAGCTCCTTCCGCCTTGCCCTTTTAATTGTGGCGGTTCTGGCG
MAACPQLPPCRVSCGGSG	1343	ATGGCAGCATGCCCTTAGCTCCCTCCCTGCCGCGTGTCTTGTGGCGGTTCTGGCG
MAACPLLPPCADDCGGSG	853	ATGGCAGCATGCCCCTTGCTCCCTCCCTGCGCTGATGATTGTGGCGGTTCTGGCG
MAACAILPPCDQNCGGSG	684	ATGGCAGCATGCGCCATTCTCCCCCCGTGCGACTAGAATTGTGGCGGTTCTGGCG
MAACLQLPPCNVSCGGSG	627	ATGGCAGCATGCCTGTAGCTTCCTCCTTGCAACGTGTCCTGTGGCGGTTCTGGCG
MAACPLLPPCSLDCGGSG	601	ATGGCAGCATGCCCCCTGCTGCCGCCTTGCAGTCTGGATTGTGGCGGTTCTGGCG
MAACPYLPPCQLACGGSG	598	ATGGCAGCATGCCCCTACCTTCCTCCCTGCTAGCTGGCGTGTGGCGGTTCTGGCG
MAACLQLPPCSSPCGGSG	548	ATGGCAGCATGCTTGTAGCTTCCTCCCTGCTCGTCCCCGTGTGGCGGTTCTGGCG
MAACPQLPPCGTFCGGSG	461	ATGGCAGCATGCCCTTAGCTGCCCCCTTGCGGTACTTTTTGTGGCGGTTCTGGCG
MAACPALPPCQLSCGGSG	447	ATGGCAGCATGCCCTGCGTTGCCCCCGTGCTAGTTGTCTTGTGGCGGTTCTGGCG
MAACPQLPPCLYPCGGSG	400	ATGGCAGCATGCCCCTAGCTCCCCCCGTGCCTGTATCCTTGTGGCGGTTCTGGCG
MAACQQGCTVLPPCGGSG	383	ATGGCAGCATGCTAGCAGGGCTGCACTGTGCTTCCGCCCTGTGGCGGTTCTGGCG
MAACPSLPPCPWNCGGSG	347	ATGGCAGCATGCCCGAGTCTTCCCCCGTGCCCCTGGAATTGTGGCGGTTCTGGCG
MAACSQLPPCARGCGGSG	338	ATGGCAGCATGCAGCTAGTTGCCTCCTTGCGCTCGGGGTTGTGGCGGTTCTGGCG
MAACLQLPPCNHHCGGSG	296	ATGGCAGCATGCCTCTAGCTGCCTCCTTGCAACCATCACTGTGGCGGTTCTGGCG
MAACILPPCSYTQCGGSG	292	ATGGCAGCATGCATCTTGCCTCCCTGCTCTTACACGTAGTGTGGCGGTTCTGGCG
MAACAILPPCQPRCGGSG	289	ATGGCAGCATGCGCCATTCTCCCTCCGTGCTAGCCTCGGTGTGGCGGTTCTGGCG
MAACPLLPPCGIGCGGSG	289	ATGGCAGCATGCCCTCTGCTGCCTCCGTGCGGTATTGGCTGTGGCGGTTCTGGCG
MAACYLLPPCQLGCGGSG	278	ATGGCAGCATGCTATCTGCTTCCCCCTTGCTAGTTGGGCTGTGGCGGTTCTGGCG
MAACRGTCPVLPPCGGSG	269	ATGGCAGCATGCAGGGGCACCTGCCCCCGTTCTGCCTCCTTGTGGCGGTTCTGGCG
MAVRCPLPPYQCCGGSG	262	ATGGCAGTGCGCTGCCCCTTGCCGCCGTATTAGTGTTGTGGCGGTTCTGGCG
MAACPYLPPCGESCGGSG	250	ATGGCAGCATGCCCGTACCTTCCCCCGTGCGGGGAGAGTTGTGGCGGTTCTGGCG
MAACSILPPCSQNCGGSG	246	ATGGCAGCATGCTCTATCCTTCCTCCTTGCTCTTAGAATTGTGGCGGTTCTGGCG
MAACQTGCPILPPCGGSG	246	ATGGCAGCATGCTAGACGGGTTGCCCGATTCTGCCTCCCTGTGGCGGTTCTGGCG
MAACQILPPCFQPCGGSG	242	ATGGCAGCATGCTAGATTTTGCCGCCGTGCTTTCAGCCCTGTGGCGGTTCTGGCG
MAACPSLPPCNQHCGGSG	242	ATGGCAGCATGCCCCTCTCTCCCTCCCTGCAATTAGCATTGTGGCGGTTCTGGCG
MAACPYLPPCPLDCGGSG	210	ATGGCAGCATGCCCTTATCTCCCTCCTTGCCCTTTGGATTGTGGCGGTTCTGGCG
MAACQVLPPCGFICGGSG	202	ATGGCAGCATGCTAGGTGCTTCCTCCGTGCGGTTTCATCTGTGGCGGTTCTGGCG
MAACSLLPPCQLSCGGSG	202	ATGGCAGCATGCAGTCTGTTGCCCCCTTGCTAGTTGTCGTGTGGCGGTTCTGGCG
MAACSILPPCFQTCGGSG	190	ATGGCAGCATGCAGCATCTTGCCCCCCTGCTTCTAGACCTGTGGCGGTTCTGGCG
MAARMKSSCLPPCCGGSG	185	ATGGCAGCACGCATGAAGAGTAGTTGCCTCCCTCCGTGCTGTGGCGGTTCTGGCG
MAACAQLPPCSLPCGGSG	183	ATGGCAGCATGCGCTTAGTTGCCGCCGTGCTCTCTTCCGTGTGGCGGTTCTGGCG
MAACYQLPPCDHSCGGSG	177	ATGGCAGCATGCTACTAGTTGCCTCCCTGCGATCACAGTTGTGGCGGTTCTGGCG
MAACPQLPPCVLACGGSG	171	ATGGCAGCATGCCCTTAGCTTCCGCCGTGCGTGCTCGCGTGTGGCGGTTCTGGCG
MAACKRTHCLPPCCGGSG	169	ATGGCAGCATGCAAGCGTACCCATTGCCTCCCCCCCTGTTGTGGCGGTTCTGGCG
MAACRQLPPCSDPCGGSG	164	ATGGCAGCATGCCGTTAGTTGCCTCCCTGCAGCGATCCCTGTGGCGGTTCTGGCG
MAACRGHCPILPPCGGSG	162	ATGGCAGCATGCCGTGGCCATTGCCCTATCCTCCCTCCTTGTGGCGGTTCTGGCG
MAACYLPPCQLQLCGGSG	157	ATGGCAGCATGCTATCTTCCCCCTTGCCAGCTGTAGTTGTGTGGCGGTTCTGGCG
MAACSILPPCTTHCGGSG	154	ATGGCAGCATGCTCCATTTTGCCGCCTTGCACGACGCATTGTGGCGGTTCTGGCG
MAACSRHCLTLPPCGGSG	149	ATGGCAGCATGCAGTCGTCATTGCCTGACTCTTCCTCCGTGTGGCGGTTCTGGCG
MAACPVLPPCSRPCGGSG	149	ATGGCAGCATGCCCCGTGTTGCCCCCTTGCAGTCGTCCTTGTGGCGGTTCTGGCG
MAACLQLPPCDFQCGGSG	147	ATGGCAGCATGCCTCTAGCTTCCTCCCTGCGATTTTCAGTGTGGCGGTTCTGGCG
MAACPYLPPCGTICGGSG	147	ATGGCAGCATGCCCCTATCTGCCTCCTTGCGGCACGATCTGTGGCGGTTCTGGCG
MAACALLPPCNQNCGGSG	145	ATGGCAGCATGCGCTCTGCTGCCGCCCTGCAATTAGAACTGTGGCGGTTCTGGCG
MAACHSRCPQLPPCGGSG	138	ATGGCAGCATGCCATAGTAGGTGCCCCTAGCTTCCCCCGTGTGGCGGTTCTGGCG
MAACLPPCPLLPCGGSG	133	ATGGCAGCATGCCTGCCTCCTTGCCCGTTGTTGCCGTGTGGCGGTTCTGGCG
MAACKLLPPCQFECGGSG	130	ATGGCAGCATGCAAGCTTTTGCCTCCGTGCTAGTTCGAGTGTGGCGGTTCTGGCG
MAACYGQCTQLPPCGGSG	129	ATGGCAGCATGCTACGGTCAGTGCACTTAGCTGCCTCCTTGTGGCGGTTCTGGCG
MAACPYLPPCSYTCGGSG	129	ATGGCAGCATGCCCTTACCTTCCGCCGTGCTCCTACACTTGTGGCGGTTCTGGCG
MAACPFLPPCSSACGGSG	126	ATGGCAGCATGCCCTTTCTTGCCCCCCTGCTCGAGTGCTTGTGGCGGTTCTGGCG
MAACGQLPPCDAMCGGSG	126	ATGGCAGCATGCGGCTAGCTGCCCCCCTGCGATGCTATGTGTGGCGGTTCTGGCG
MAACQTYASCVLRCGGSG	125	ATGGCAGCATGCTAGACTTATGCCTCGTGCGTGCTCCGGTGTGGCGGTTCTGGCG
MAACGYLPPCPLSCGGSG	122	ATGGCAGCATGCGGTTATCTCCCTCCCTGCCCTCTCTCCTGTGGCGGTTCTGGCG
MAACSVLPPCSFSCGGSG	122	ATGGCAGCATGCTCTGTTCTTCCTCCCTGCTCCTTCTCGTGTGGCGGTTCTGGCG
MAACILPPCPSSCGGSG	122	ATGGCAGCATGCATTCTCCCCCCGTGCCCTTCTTCGTGTGGCGGTTCTGGCG
MAACPVLPPCPINCGGSG	122	ATGGCAGCATGCCCCGTCCTGCCTCCCTGCCCTATTAACTGTGGCGGTTCTGGCG
MAACSRSCPVLPPCGGSG	121	ATGGCAGCATGCTCTCGTTCCTGCCCTGTGCTTCCCCCCTGTGGCGGTTCTGGCG
MAACYGECQVLPPCGGSG	116	ATGGCAGCATGCTACGGCGAGTGCTAGGTGTTGCCTCCTTGTGGCGGTTCTGGCG
MAACPSQCTILPPCGGSG	112	ATGGCAGCATGCCCCTCCTAGTGCACCATTCTTCCGCCCTGTGGCGGTTCTGGCG
MAACGVLPPCQRECGGSG	107	ATGGCAGCATGCGGCGTTCTTCCTCCCTGCTAGAGGGAGTGTGGCGGTTCTGGCG
MAACTSRCPQLPPCGGSG	105	ATGGCAGCATGCACCTCGCGGTGCCCTCAGCTCCCTCCGTGTGGCGGTTCTGGCG
MAACQLLPPCSYSCGGSG	104	ATGGCAGCATGCTAGCTGTTGCCCCCGTGCTCTTATTCCTGTGGCGGTTCTGGCG
MAACNTKCSILPPCGGSG	101	ATGGCAGCATGCAACACTAAGTGCTCTATTCTCCCCCCCTGTGGCGGTTCTGGCG
MAACTVLPPCNYKCGGSG	98	ATGGCAGCATGCACCGTCTTGCCCCCGTGCAATTATAAGTGTGGCGGTTCTGGCG
MAACFDRCLQLPPCGGSG	97	ATGGCAGCATGCTTCGATCGCTGCCTTTAGTTGCCCCCTTGTGGCGGTTCTGGCG
MAACHSRCPTLPPCGGSG	97	ATGGCAGCATGCCACTCTCGCTGCCCCACTTTGCCCCCTTGTGGCGGTTCTGGCG
MAACLQLPPCAWTCGGSG	93	ATGGCAGCATGCTTGCAGCTGCCCCCTTGCGCTTGGACGTGTGGCGGTTCTGGCG
MAACPVLPPCISNCGGSG	91	ATGGCAGCATGCCCCGTTCTCCCCCCCTGCATTAGTAATTGTGGCGGTTCTGGCG

MAACSILPPCTIQCGGSG	89	ATGGCAGCATGCTCTATTCTTCCTCCGTGCACGATTTAGTGTGGCGGTTCTGGCG
MAACPYLPPCDFLCGGSG	87	ATGGCAGCATGCCCGTATCTCCCCCCCTGCGACTTCCTCTGTGGCGGTTCTGGCG
MAACSGYCPYLPPCGGSG	86	ATGGCAGCATGCAGCGGTTATTGCCCCTACCTCCCGCCTTGTGGCGGTTCTGGCG
MAACPVLPPCRSDCGGSG	86	ATGGCAGCATGCCCCGTGTTGCCGCCTTGCCGCTCGGATTGTGGCGGTTCTGGCG
MAACPSLPPCRESCGGSG	84	ATGGCAGCATGCCCGAGCCTTCCCCCTTGCCGTGAGAGCTGTGGCGGTTCTGGCG
MAACSNRCTLLPPCGGSG	83	ATGGCAGCATGCTCTAATCGGTGCACCTTGTTGCCGCCTTGTGGCGGTTCTGGCG
MAACSILPPCNSPCGGSG	83	ATGGCAGCATGCTCGATTCTGCCGCCGTGCAATAGTCCTTGTGGCGGTTCTGGCG
MAACLQLPPCSLSCGGSG	82	ATGGCAGCATGCTTGTAGCTCCCTCCGTGCAGTCTCAGCTGTGGCGGTTCTGGCG
MAACVSTCQILPPCGGSG	82	ATGGCAGCATGCGTTAGTACGTGCTAGATCCTCCCCCCTTGTGGCGGTTCTGGCG
MAACNVLPPCSSHCGGSG	81	ATGGCAGCATGCAACGTCTTGCCGCCTTGCTCCTCTCATTGTGGCGGTTCTGGCG
MAACNIQSCLPPCGGSG	78	ATGGCAGCATGCAACATTTAGTCTTGCCTTCCTCCTTGTGGCGGTTCTGGCG
MAACSQGCPSLPPCGGSG	77	ATGGCAGCATGCAGTTAGGGGTGCCCGTCGCTGCCTCCGTGTGGCGGTTCTGGCG
MAACPFLPPCSMSCGGSG	76	ATGGCAGCATGCCCTTTTCTGCCTCCGTGCTCTATGTCCTGTGGCGGTTCTGGCG
MAACSILPPCRSGCGGSG	74	ATGGCAGCATGCTCCATTCTCCCTCCTTGCCGCAGCGGGTGTGGCGGTTCTGGCG
MAACPHLPPCNTLCGGSG	74	ATGGCAGCATGCCCTCATCTTCCCCCTTGCAACACTTTGTGTGGCGGTTCTGGCG
MAACAAYCPQLPPCGGSG	73	ATGGCAGCATGCGCTGCTTATTGCCCCTAGCTTCCTCCTTGTGGCGGTTCTGGCG
MAACQSGCGILPPCGGSG	72	ATGGCAGCATGCTAGAGCGGTTGCGGCATCTTGCCTCCTTGTGGCGGTTCTGGCG
MAACAVGCPILPPCGGSG	72	ATGGCAGCATGCGCCGTTGGGTGCCCCATTTTGCCTCCCTGTGGCGGTTCTGGCG
MAACPYLPPCNMQCGGSG	71	ATGGCAGCATGCCCCTACCTCCCGCCTTGCAATATGTAGTGTGGCGGTTCTGGCG
MAACSQKCFQLPPCGGSG	70	ATGGCAGCATGCTCTTAGAAGTGCTTTTAGCTTCCTCCTTGTGGCGGTTCTGGCG
MAACILPPCPFSCGGSG	68	ATGGCAGCATGCATCTTGCCGCCTTGCCCCTTTTCCTGTGGCGGTTCTGGCG
MAACLRNCPVLPPCGGSG	68	ATGGCAGCATGCTTGAGGAATTGCCCTGTTCTTCCTCCTTGTGGCGGTTCTGGCG
MAACSVRCGILPPCGGSG	67	ATGGCAGCATGCAGTGTTCGCTGCGGTATTTTGCCTCCTTGTGGCGGTTCTGGCG
MAACPLLPPCGANCGGSG	67	ATGGCAGCATGCCCTCTTCTGCCCCCTTGCGGTGCGAATTGTGGCGGTTCTGGCG
MAACRNQCLILPPCGGSG	67	ATGGCAGCATGCCGCAACCAGTGCTTGATTCTTCCTCCCTGTGGCGGTTCTGGCG
MAACAILPPCTLTCGGSG	65	ATGGCAGCATGCGCCATTCTCCCCCCTTGCACGCTTACTTGTGGCGGTTCTGGCG
MAACILPPCQFKDCGGSG	64	ATGGCAGCATGCATTCTGCCTCCCTGCTAGTTCAAGGATTGTGGCGGTTCTGGCG
MAACGQLPPCSVVCGGSG	64	ATGGCAGCATGCGGGTAGCTCCCTCCCTGCTCCGTTGTCTGTGGCGGTTCTGGCG
MAACNTLCPYLPPCGGSG	63	ATGGCAGCATGCAATACTCTTTGCCCTTATCTGCCCCCTTGTGGCGGTTCTGGCG
MAACVGRCEVLPPCGGSG	62	ATGGCAGCATGCGTCGGGCGTTGCGAGGTTTTGCCTCCTTGTGGCGGTTCTGGCG
MAACPQLPPCHVYCGGSG	62	ATGGCAGCATGCCCTTAGCTCCCTCCGTGCCACGTTTACTGTGGCGGTTCTGGCG
MAACALLPPCPNACGGSG	62	ATGGCAGCATGCGCGCTGTTGCCGCCTTGCCCTAATGCGTGTGGCGGTTCTGGCG
MAACQVRCDILPPCGGSG	62	ATGGCAGCATGCTAGGTCCGGTGCGACATCCTCCCGCCTTGTGGCGGTTCTGGCG
MAACQFPCLVLPPCGGSG	62	ATGGCAGCATGCTAGTTTCCCTGCCTCGTCCTCCCTCCGTGTGGCGGTTCTGGCG

Library B

MAACSILPPCNPPQCGGSG	4513	ATGGCAGCATGCTCGATTCTTCCTCCGTGCAATCCTCCGTAGTGTGGCGGTTCTGGCG
MAACPLLPPCHLPQCGGSG	2041	ATGGCAGCATGCCCGTTGCTTCCGCCGTGCCATCTTCCTTAGTGTGGCGGTTCTGGCG
MAACQVLPPCGLQLCGGSG	1877	ATGGCAGCATGCCAGGTGTTGCCTCCGTGCGGTCTGTAGCTGTGTGGCGGTTCTGGCG
MAACRQLPPCAEYVCGGSG	1291	ATGGCAGCATGCCGGTAGCTGCCTCCTTGCGCTGAGTATGTTTGTGGCGGTTCTGGCG
MAACPELPPCQLMLCGGSG	1048	ATGGCAGCATGCCCGGAGCTTCCGCCGTGCTAGTTGATGCTGTGTGGCGGTTCTGGCG
MAACPMLPPCDLSYCGGSG	650	ATGGCAGCATGCCCTATGCTGCCTCCGTGCGATCTGAGTTATTGTGGCGGTTCTGGCG
MAACGLLPPCHQFHCGGSG	600	ATGGCAGCATGCGGGCTTTTGCCGCCTTGCCATTAGTTTCATTGTGGCGGTTCTGGCG
MAACTLLPPCTPDQCGGSG	580	ATGGCAGCATGCACGCTGTTGCCTCCTTGCACGCCGGATTAGTGTGGCGGTTCTGGCG
MAACRPKQCWQLPPCGGSG	571	ATGGCAGCATGCCGTCCGAAGCAGTGCTGGCAGTTGCCTCCGTGTGGCGGTTCTGGCG
MAACAQLPPCDYSGCGGSG	459	ATGGCAGCATGCGCGTAGCTTCCTCCGTGCGATTATTCTGGGTGTGGCGGTTCTGGCG
MAACQILPPCHSPGCGGSG	413	ATGGCAGCATGCTAGATTCTGCCTCCTTGCCATTCGCCGGGGTGTGGCGGTTCTGGCG
MAACPSLPPCWQLQCGGSG	244	ATGGCAGCATGCCCTAGTTTGCCTCCTTGCTGGCAGTTGTAGTGTGGCGGTTCTGGCG
MAACEILPPCLQFQCGGSG	224	ATGGCAGCATGCGAGATTTTGCCTCCTTGCCTTTAGTTTCAGTGTGGCGGTTCTGGCG
MAACVWNNCSLLPPCGGSG	201	ATGGCAGCATGCGTGTGGAATAATTGCTCTCTTCTGCCTCCTTGTGGCGGTTCTGGCG
MAACPYLPPCSWDLCGGSG	192	ATGGCAGCATGCCCTTATCTTCCTCCGTGCTCGTGGGATCTGTGTGGCGGTTCTGGCG
MAACLELPPCQFASCGGSG	183	ATGGCAGCATGCTTGGAGTTGCCGCCGTGCCAGTTTGCTTCGTGTGGCGGTTCTGGCG
MAACPELPPCTVLKCGGSG	176	ATGGCAGCATGCCCTGAGCTGCCGCCTTGCACGGTTCTTAAATGTGGCGGTTCTGGCG
MAACIQLPPCQQSSCGGSG	175	ATGGCAGCATGCATTCAGCTTCCGCCTTGCTAGCAGTCTTCGTGTGGCGGTTCTGGCG
MAACQLLPPCQFLQCGGSG	169	ATGGCAGCATGCCAGCTTCTGCCTCCTTGCTAGTTTTTGCAGTGTGGCGGTTCTGGCG
MAACLPPHSCWNQVCGGSG	161	ATGGCAGCATGCCTGCCTCCGCATTCTTGCTGGAATCAGGTTTGTGGCGGTTCTGGCG
MAACSQLPPCTYLSCGGSG	150	ATGGCAGCATGCTCTCAGCTTCCGCCTTGCACTTATCTTTCGTGTGGCGGTTCTGGCG
MAACRYLPPCPYKLCGGSG	149	ATGGCAGCATGCCGGTATCTTCCTCCTTGCCCTTATAAGCTGTGTGGCGGTTCTGGCG
MAACLYPRCPSLPPCGGSG	143	ATGGCAGCATGCCTGTATCCGCGTTGCCCTTCTCTTCCTCCGTGTGGCGGTTCTGGCG
MAACLQLPPCGVSLCGGSG	137	ATGGCAGCATGCCTGTAGCTGCCGCCTTGCGGTGTTAGTCTGTGTGGCGGTTCTGGCG
MAACQLLPPCAIQWCGGSG	126	ATGGCAGCATGCCAGTTGCTTCCGCCGTGCGCGATTTAGTGGTGTGGCGGTTCTGGCG
MAACAILPPCGQLSCGGSG	123	ATGGCAGCATGCGCTATTCTGCCGCCGTGCGGGCAGCTTAGTTGTGGCGGTTCTGGCG
MAACLTLPPCQAVSCGGSG	122	ATGGCAGCATGCCTTACTCTTCCTCCTTGCTAGGCGGTTTCTTGTGGCGGTTCTGGCG
MAACYLLPPCPSTSCGGSG	122	ATGGCAGCATGCTATTTGTTGCCGCCGTGCCCTTCGACGTCTTGTGGCGGTTCTGGCG
MAACLTLPPCPSFTCGGSG	106	ATGGCAGCATGCTTGACTTTGCCGCCGTGCCCGTCTTTTACTTGTGGCGGTTCTGGCG
MAACILPPCPTSEYCGGSG	104	ATGGCAGCATGCATTCTGCCTCCTTGCCCGACTTCGGAGTATTGTGGCGGTTCTGGCG
MAACSVLPPCSFVACGGSG	100	ATGGCAGCATGCTCGGTGTTGCCTCCTTGCTCTTTTGTTGCGTGTGGCGGTTCTGGCG
MAACHYLPPCQPAICGGSG	90	ATGGCAGCATGCCATTATTTGCCGCCTTGCTAGCCGGCGATTTGTGGCGGTTCTGGCG
MAACSLLPPCHSTNCGGSG	73	ATGGCAGCATGCAGTCTTCTTCCTCCGTGCCATTCGACGAATTGTGGCGGTTCTGGCG
MAACPLLPPCHLSFCGGSG	70	ATGGCAGCATGCCCGTTGCTGCCTCCGTGCCATCTTTCTTTTTGTGGCGGTTCTGGCG
MAACQSTFCPILPPCGGSG	68	ATGGCAGCATGCTAGTCTACGTTTTGCCCGATTCTGCCTCCTTGTGGCGGTTCTGGCG
MAACGYLPPCNYLHCGGSG	61	ATGGCAGCATGCGGTTATTTGCCTCCTTGCAATTATCTGCATTGTGGCGGTTCTGGCG
MAACTLLPPCSSLQCGGSG	59	ATGGCAGCATGCACTTTGTTGCCGCCGTGCTCTTCTCTTCAGTGTGGCGGTTCTGGCG
MAACPLLPPCRNTVCGGSG	57	ATGGCAGCATGCCCTCTGTTGCCTCCTTGCCGGAATACGGTTTGTGGCGGTTCTGGCG
MAACPNQTCPLLPPCGGSG	53	ATGGCAGCATGCCCTAATTAGACGTGCCCTTTGCTGCCTCCTTGTGGCGGTTCTGGCG

MAACKVLPPCSAQRCGGSG	52	ATGGCAGCATGCAAGGTTCTGCCTCCGTGCTCGGCTTAGCGGTGTGGCGGTTCTGGCG
MAACLLLPPCPIMTCGGSG	48	ATGGCAGCATGCCTTCTGCTTCCGCCTTGCCCGATTATGACGTGTGGCGGTTCTGGCG
MAACLLASCPILPPCGGSG	47	ATGGCAGCATGCTTGCTGGCTTCGTGCCCGATTCTTCCTCCGTGTGGCGGTTCTGGCG
MAACILPPCPTTIPCGGSG	43	ATGGCAGCATGCATTCTTCCGCCGTGCCCGACGACGATTCCGTGTGGCGGTTCTGGCG
MAACRHLPPCTQLRCGGSG	42	ATGGCAGCATGCCGTCATCTGCCGCCGTGCACGTAGCTTCGGTGTGGCGGTTCTGGCG
MAACAQLPPCPSQACGGSG	40	ATGGCAGCATGCGCGTAGTTGCCTCCGTGCCCTTCTTAGGCGTGTGGCGGTTCTGGCG
MAACSALPPCQWAQCGGSG	40	ATGGCAGCATGCAGTGCTCTGCCGCCTTGCTAGTGGGCTCAGTGTGGCGGTTCTGGCG
MAACHLLPPCGRQSCGGSG	37	ATGGCAGCATGCCATCTTTTGCCTCCGTGCGGTCGTTAGTCGTGTGGCGGTTCTGGCG
MAACSRLPPCVILDCGGSG	36	ATGGCAGCATGCTCTAGGCTTCCTCCTTGCGTTATTCTTGATTGTGGCGGTTCTGGCG
MAACQQLPPCMPRFCGGSG	36	ATGGCAGCATGCCAGTAGTTGCCTCCTTGCATGCCTCGGTTTTGTGGCGGTTCTGGCG
MAACPVLPPCFTRLCGGSG	35	ATGGCAGCATGCCCGGTTCTGCCTCCGTGCTTTACTCGTTTGTGTGGCGGTTCTGGCG
MAACSLLPPCTFVACGGSG	33	ATGGCAGCATGCTCGTTGCTGCCGCCGTGCACTTTTGTGGCGTGTGGCGGTTCTGGCG
MAACAILPPCHFRSCGGSG	30	ATGGCAGCATGCGCGATTCTTCCGCCTTGCCATTTTCGGTCGTGTGGCGGTTCTGGCG
MAACYLLPPCAISVCGGSG	29	ATGGCAGCATGCTATCTTTTGCCGCCTTGCGCGATTTCGGTTTGTGGCGGTTCTGGCG
MAACALLPPCFIQECGGSG	29	ATGGCAGCATGCGCTCTGTTGCCGCCGTGCTTTATTTAGGAGTGTGGCGGTTCTGGCG
MAACLLLPPCPSINCGGSG	29	ATGGCAGCATGCCTTCTGCTTCCTCCGTGCCCTTCGATTAATTGTGGCGGTTCTGGCG
MAACLTLPPCVLQQCGGSG	28	ATGGCAGCATGCTTGACGCTGCCGCCGTGCGTGCTGCAGTAGTGTGGCGGTTCTGGCG
MAACPQLPPCIHSYCGGSG	28	ATGGCAGCATGCCCTCAGCTGCCTCCGTGCATTCATTCTTATTGTGGCGGTTCTGGCG
MAACKFLPPCGGSG	27	ATGGCAGCATGCAAGTTTCTTCCTCCTTGTGGCGGTTCTGGCG
MAACPMLPPCGIFPCGGSG	27	ATGGCAGCATGCCCGATGCTTCCGCCTTGCGGGATTTTTCCTTGTGGCGGTTCTGGCG
MAACRVLPPCSEYNCGGSG	26	ATGGCAGCATGCAGGGTGTTGCCTCCGTGCTCTGAGTATAATTGTGGCGGTTCTGGCG
MAACPILPPCPPTQCGGSG	26	ATGGCAGCATGCCCTATTTTGCCTCCTTGCCCGCCGACTCAGTGTGGCGGTTCTGGCG
MAACLLLPPCAPTTCGGSG	26	ATGGCAGCATGCTTGTTGCTGCCGCCTTGCGCGCCGACTACTTGTGGCGGTTCTGGCG
MAACSLLPPCTPVLCGGSG	26	ATGGCAGCATGCTCTCTGCTTCCGCCTTGCACTCCTGTTTTGTGTGGCGGTTCTGGCG
MAACVLPPCHWSLQCGGSG	26	ATGGCAGCATGCGTTCTTCCTCCTTGCCATTGGTCGCTGCAGTGTGGCGGTTCTGGCG
MAACVVLPPCPLQTCGGSG	24	ATGGCAGCATGCGTGGTGTTGCCTCCGTGCCCGTTGCAGACTTGTGGCGGTTCTGGCG
MAACNLLPPCPPLSCGGSG	22	ATGGCAGCATGCAATCTGCTGCCTCCTTGCCCGCCTCTGTCTTGTGGCGGTTCTGGCG
MAACSTLPPCQLLHCGGSG	21	ATGGCAGCATGCAGTACGCTGCCGCCTTGCTAGTTGTTGCATTGTGGCGGTTCTGGCG
MAACQLGSCMVLPPCGGSG	21	ATGGCAGCATGCTAGCTTGGTTCTTGCATGGTGTTGCCTCCGTGTGGCGGTTCTGGCG
MAACARSTCPQLPPCGGSG	21	ATGGCAGCATGCGCGCGGTCTACGTGCCCGCAGCTGCCTCCTTGTGGCGGTTCTGGCG
MAACQYLHCSVLPPCGGSG	20	ATGGCAGCATGCTAGTATTTGCATTGCTCTGTTCTGCCTCCGTGTGGCGGTTCTGGCG
MAACFLLPPCQVALCGGSG	20	ATGGCAGCATGCTTTCTTTTGCCTCCTTGCTAGGTTGCGCTTTGTGGCGGTTCTGGCG
MAACPTLPPCHNQLCGGSG	19	ATGGCAGCATGCCCGACGCTTCCTCCGTGCCATAATTAGTTGTGTGGCGGTTCTGGCG
MAACMVLPPCPHQRCGGSG	19	ATGGCAGCATGCATGGTGTTGCCTCCGTGCCCTCATCAGCGGTGTGGCGGTTCTGGCG
MAACDAWRCPVLPPCGGSG	19	ATGGCAGCATGCGATGCTTGGAGGTGCCCTGTGTTGCCTCCGTGTGGCGGTTCTGGCG
MAACSLLPPCGSTTCGGSG	18	ATGGCAGCATGCTCTTTGTTGCCTCCTTGCGGTTCTACTACGTGTGGCGGTTCTGGCG
MAACPVLPPCRTIWCGGSG	18	ATGGCAGCATGCCCGGTGCTGCCTCCGTGCCGTACTATTTGGTGTGGCGGTTCTGGCG
MAACPPQPCPYLPPCGGSG	17	ATGGCAGCATGCCCGCCTTAGCCGTGCCCTTATCTGCCTCCTTGTGGCGGTTCTGGCG
MAACPTLPPCQPGYCGGSG	16	ATGGCAGCATGCCCGACGTTGCCTCCTTGCCAGCCTGGTTATTGTGGCGGTTCTGGCG
MAACLPPHNCPLSLCGGSG	16	ATGGCAGCATGCCTGCCGCCTCATAATTGCCCGCTTTCTCTTTGTGGCGGTTCTGGCG
MAACFTAKCLQLPPCGGSG	16	ATGGCAGCATGCTTTACGGCTAAGTGCCTTTAGCTTCCGCCTTGTGGCGGTTCTGGCG
MAACRFLSCSVLPPCGGSG	16	ATGGCAGCATGCAGGTTTTTGAGTTGCAGTGTTTTGCCTCCTTGTGGCGGTTCTGGCG
MAACYHFTCPVLPPCGGSG	14	ATGGCAGCATGCTATCATTTTACTTGCCCTGTGCTTCCTCCTTGTGGCGGTTCTGGCG
MAACSRVPCNQLPPCGGSG	13	ATGGCAGCATGCTCTCGGGTTCCTTGCAATTAGCTTCCTCCGTGTGGCGGTTCTGGCG
MAACQELPPCLQIQCGGSG	12	ATGGCAGCATGCCAGGAGCTTCCGCCTTGCCTGCAGATTTAGTGTGGCGGTTCTGGCG
MAACPTKSCLLLPPCGGSG	12	ATGGCAGCATGCCCTACGAAGTCTTGCCTTTTGCTTCCTCCGTGTGGCGGTTCTGGCG
MAACPTARCTQLPPCGGSG	12	ATGGCAGCATGCCCTACGGCGCGGTGCACGTAGTTGCCTCCTTGTGGCGGTTCTGGCG
MAACFYLPPCQDLFCGGSG	12	ATGGCAGCATGCTTTTATCTGCCTCCGTGCTAGGATTTGTTTTGTGGCGGTTCTGGCG
MAACYLPPCPSLPHCGGSG	11	ATGGCAGCATGCTATCTTCCTCCGTGCCCGTCTCTGCCTCATTGTGGCGGTTCTGGCG
MAACLRRPCDILPPCGGSG	11	ATGGCAGCATGCTTGCGTCGGCCGTGCGATATTCTTCCGCCTTGTGGCGGTTCTGGCG
MAACSSLPPCYPFQCGGSG	11	ATGGCAGCATGCAGTTCGTTGCCTCCTTGCTATCCGTTTTAGTGTGGCGGTTCTGGCG
MAACGAKYCSYLPPCGGSG	11	ATGGCAGCATGCGGGGCGAAGTATTGCTCGTATTTGCCTCCTTGTGGCGGTTCTGGCG
MAACSSLPPCSTSECGGSG	11	ATGGCAGCATGCTCGAGTTTGCCTCCTTGCTCGACTTCTGAGTGTGGCGGTTCTGGCG
MAACQYLPPCSFLVCGGSG	11	ATGGCAGCATGCCAGTATCTTCCTCCTTGCTCTTTTTTGGTTTGTGGCGGTTCTGGCG
MAACVPLRCTLLPPCGGSG	10	ATGGCAGCATGCGTTCCTCTTCGGTGCACTCTTCTGCCTCCGTGTGGCGGTTCTGGCG
MAACLLLPPCSIQGCGGSG	10	ATGGCAGCATGCCTTTTGTTGCCGCCTTGCTCTATTTAGGGTTGTGGCGGTTCTGGCG
MAACIGSLCPVLPPCGGSG	9	ATGGCAGCATGCATTGGTAGTTTGTGCCCGGTTCTTCCTCCGTGTGGCGGTTCTGGCG
MAACRTLPPCSDLTCGGSG	9	ATGGCAGCATGCCGTACGCTGCCTCCTTGCTCTGATCTTACTTGTGGCGGTTCTGGCG
MAACSQLPPCSIIKCGGSG	9	ATGGCAGCATGCTCGTAGCTTCCGCCGTGCAGTATTATTAAGTGTGGCGGTTCTGGCG
MAACLLLPPCQHQMCGGSG	8	ATGGCAGCATGCCTTCTGCTGCCGCCTTGCCAGCATTAGATGTGTGGCGGTTCTGGCG
MAACALLPPCSWVSCGGSG	8	ATGGCAGCATGCGCGTTGCTGCCGCCGTGCTCGTGGGTGTCTTGTGGCGGTTCTGGCG
MAACKTLPPCPLRLCGGSG	8	ATGGCAGCATGCAAGACTCTGCCTCCTTGCCCTCTGCGTCTTTGTGGCGGTTCTGGCG
MAACYVLPPCSNVVCGGSG	8	ATGGCAGCATGCTATGTTCTTCCTCCGTGCTCTAATGTGGTGTGTGGCGGTTCTGGCG
MAACKHLPPCSVAYCGGSG	8	ATGGCAGCATGCAAGCATTTGCCTCCTTGCTCTGTTGCTTATTGTGGCGGTTCTGGCG
MAACLVLPPCFLVDCGGSG	8	ATGGCAGCATGCCTTGTTCTTCCGCCTTGCTTTTTGGTGGATTGTGGCGGTTCTGGCG

Library 6x6

MAACRAKQVCCLAADRTCGGSG	4773	ATGGCAGCGTGTCGTGCGAAGCAGGTGTGTTGCTTGGCGGCGGATCGTACGTGTGGCGGTTCTGGCG
MAACWKRQVCCLSQDRLCGGSG	3992	ATGGCAGCGTGTTGGAAGAGGCAGGTGTGTTGCCTTTCGTAGGATAGGTTGTGTGGCGGTTCTGGCG
MAACQGILPPCGRGGGGCGGSG	3402	ATGGCAGCGTGTTAGGGGATTTTGCCGCCTTGCGGGCGTGGTGGTGGTGGGTGTGGCGGTTCTGGCG
MAACTCRKVCCLAQEMPCGGSG	2782	ATGGCAGCGTGTACGTGTCGGAAGGTGTGTTGCTTGGCGTAGGAGATGCCTTGTGGCGGTTCTGGCG
MAACTHRKGCCVAQEVVCGGSG	1709	ATGGCAGCGTGTACTCATAGGAAGGGGTGTTGCGTGGCTTAGGAGGTTGTTTGTGGCGGTTCTGGCG
MAACGPRKVCCLSQEVICGGSG	1529	ATGGCAGCGTGTGGGCCGAGGAAGGTTTGTTGCCTGAGTTAGGAGGTGATTTGTGGCGGTTCTGGCG
MAACSWLQSECSIISSGCGGSG	1406	ATGGCAGCGTGTTCTTGGTTGTAGAGTGAGTGCAGTATTATTAGTAGTGGGTGTGGCGGTTCTGGCG
MAACGQVLPPCEHSGASRGGSG	936	ATGGCAGCGTGTGGGTAGGTGCTTCCTCCTTGCGAGCATTCTGGGGCGTCGCGTGGCGGTTCTGGCG
MAACMYQAALCSSILPPCGGSG	924	ATGGCAGCGTGTATGTATTAGGCGGCGTTGTGCTCTAGTATTCTTCCTCCGTGTGGCGGTTCTGGCG

MAACMQRQVCCLSYEKQCGGSG	821	ATGGCAGCGTGTATGTAGAGGCAGGTGTGTTGCCTGAGTTATGAGAAGCAGTGTGGCGGTTCTGGCG
MAACWRKTVCCLAAERSCGGSG	768	ATGGCAGCGTGTTGGCGGAAGACTGTGTGTTGCCTGGCTGCTGAGAGGAGTTGTGGCGGTTCTGGCG
MAACERRAVCCVSPELHCGGSG	658	ATGGCAGCGTGTGAGAGGCGGGCTGTTTGTTGCGTTTCTCCTGAGTTGCATTGTGGCGGTTCTGGCG
MAACLQWRADCQHVLPPCGGSG	454	ATGGCAGCGTGTCTTCAGTGGAGGGCGGATTGCTAGCATGTGCTGCCTCCGTGTGGCGGTTCTGGCG
MAACYGVLPPCQFTGGPCGGSG	442	ATGGCAGCGTGTTATGGTGTTCTGCCTCCGTGCTAGTTTACTGGTGGTCCTTGTGGCGGTTCTGGCG
MAACIWAQGYCGEPARNCGGSG	421	ATGGCAGCGTGTATTTGGGCGTAGGGGTATTGCGGGGAGCCTGCTCGGAATTGTGGCGGTTCTGGCG
MAACGVSQAFSLPPWHCGGSG	402	ATGGCAGCGTGTGGCGTCTCATAGGCCTTCTCGTTGCCTCCGTGGCATTGTGGCGGTTCTGGCG
MAACQGGVVTCPHILPPCGGSG	400	ATGGCAGCGTGTTAGGGTGGGGTTGTGACGTGCCCTCATATTTTGCCGCCGTGTGGCGGTTCTGGCG
MAACFTRSVCCLSKELLCGGSG	372	ATGGCAGCGTGTTTTACTCGTAGTGTGTGTTGCCTTTCTAAGGAGCTGCTGTGTGGCGGTTCTGGCG
MAACLSRQVCCLQKDLPCGGSG	346	ATGGCAGCGTGTCTTAGTAGGCAGGTTTGTTGCTTGTAGAAGGATCTGCCTTGTGGCGGTTCTGGCG
MAACGGILPPCTYQVPACGGSG	313	ATGGCAGCGTGTGGGGGGATTCTTCCTCCTTGCACGTATTAGGTGCCTGCGTGTGGCGGTTCTGGCG
MAACQNYLKSCGLPPYDCGGSG	299	ATGGCAGCGTGTTAGAATTATTTGAAGAGTTGCGGGTTGCCTCCGTATGATTGTGGCGGTTCTGGCG
MAACALLPPQCPAYVESCGGSG	298	ATGGCAGCGTGTGCTCTGTTGCCTCCGTAGTGCCCGGCGTATGTTGAGAGTTGTGGCGGTTCTGGCG
MAACSGRYACCLSQELECGGSG	296	ATGGCAGCGTGTAGTGGGAGGTATGCGTGTTGCTTGTCGTAGGAGTTGGAGTGTGGCGGTTCTGGCG
MAACSGEYGKCQVLPPCGGSG	249	ATGGCAGCGTGTTCGGGGGAGTATGGTAAGTGCTAGGTGCTGCCTCCGTGTGGCGGTTCTGGCG
MAACHKQSNRCVGVLPPCGGSG	243	ATGGCAGCGTGTCATAAGTAGAGTAATCGGTGCGTGGGTGTTTTGCCGCCGTGTGGCGGTTCTGGCG
MAACKLRQVCCVDAAGYCGGSG	226	ATGGCAGCGTGTAAGTTGAGGTAGGTTTGTTGCGTGGATGCTGCGGGGTATTGTGGCGGTTCTGGCG
MAWCGGQLPPCGGSG	223	ATGGCTTGGTGCGGGGGTTAGTTGCCGCCTTGTGGCGGTTCTGGCG
MAACLTVLPPYSSSCGGSG	194	ATGGCAGCGTGTCTGACGGTTTTGCCGCCTTATTCTAGTAGTTGTGGCGGTTCTGGCG
MAACQDVTDGCTGILPPCGGSG	189	ATGGCAGCGTGTTAGGATGTGACTGATGGTTGCACTGGTATTCTTCCTCCGTGTGGCGGTTCTGGCG
MAACQGILPPCNVGRTGCGGSG	184	ATGGCAGCGTGTTAGGGGATTCTGCCGCCGTGCAATGTGGGTCGTACTGGTTGTGGCGGTTCTGGCG
MAACPYQLPPCDSSPLDCGGSG	183	ATGGCAGCGTGTCCGTATTAGCTTCCGCCTTGCGATTCTAGTCCTCTGGATTGTGGCGGTTCTGGCG
MAACDLVLPPCWQGSQGWGGSG	174	ATGGCAGCGTGTGATCTGGTTCTGCCGCCTTGCTGGTAGGGGTCTTAGGGGTGGGGCGGTTCTGGCG
MAACGGARKCCLDTSHSCGGSG	173	ATGGCAGCGTGTGGGGGTGCGAGGAAGTGTTGCTTGGATACGTCTCATTCTTGTGGCGGTTCTGGCG
MASCQTLLPPCGGSG	171	ATGGCTTCGTGCTAGACGCTGCTTCCTCCTTGTGGCGGTTCTGGCG
MAACGGPLKKCCGRLPPCGGSG	161	ATGGCAGCGTGTGGTGGGCCTTTGAAGAAGTGCTGTGGTAGGTTGCCTCCTTGTGGCGGTTCTGGCG
MAACGGVLPPCQLEKVECGGSG	158	ATGGCAGCGTGTGGTGGGGTGTTGCCTCCGTGCTAGTTGGAGAAGGTTGAGTGTGGCGGTTCTGGCG
MAACSWARVCCLQIDKECGGSG	155	ATGGCAGCGTGTTCTTGGGCCGCGTGTGTGTTGCTTGTAGATTGATAAGGAGTGTGGCGGTTCTGGCG
MAACLGVLPPCGMFNFQRGGSG	152	ATGGCAGCGTGTTTGGGTGTGTTGCCTCCGTGCGGTATGTTTAATTTTTAGCGTGGCGGTTCTGGCG
MAACMKKGVCCLAPDVRCGGSG	141	ATGGCAGCGTGTATGAAGAAGGGGGTGTGTTGCCTTGCTCCGGATGTGAGGTGTGGCGGTTCTGGCG
MAACTNALQRCGGTLPPCGGSG	136	ATGGCAGCGTGTACGAATGCTCTGTAGAGGTGCGGTGGTACGCTTCCGCCGTGTGGCGGTTCTGGCG
MAACVKRVALCCGQLPPCGGSG	134	ATGGCAGCGTGTGTGAAGCGTGTTGCGTTGTGCTGTGGTTAGCTTCCTCCGTGTGGCGGTTCTGGCG
MAACVGTERACFTQLPPCGGSG	130	ATGGCAGCGTGTGTGGGTACTGAGAGGGCGTGCTTTACTTAGTTGCCTCCTTGTGGCGGTTCTGGCG
MAACRWDLQECSAYLPPCGGSG	130	ATGGCAGCGTGTCGGTGGGATTTGTAGGAGTGCAGTGCGTATCTTCCTCCGTGTGGCGGTTCTGGCG
MAACDFQLQQCAFILPPCGGSG	125	ATGGCAGCGTGTGATTTTTAGTTGCAGCAGTGCGCGTTTATTCTTCCGCCTTGTGGCGGTTCTGGCG
MAACGGQSLECRIYLPPCGGSG	121	ATGGCAGCGTGTGGTGGTTAGTCTTTGGAGTGCCGGATTTATCTTCCTCCTTGTGGCGGTTCTGGCG
MAACSTTSLRCGGQLPPCGGSG	121	ATGGCAGCGTGTTCGACTACTTCTCTGAGGTGCGGTGGGTAGTTGCCTCCTTGTGGCGGTTCTGGCG
MAACQLAKRCCGGALPPCGGSG	118	ATGGCAGCGTGTTAGCTTGCTAAGCGGTGTTGCGGGGGGGCTTTGCCTCCGTGTGGCGGTTCTGGCG
MAACEWLDRACSQVLPPCGGSG	116	ATGGCAGCGTGTGAGTGGCTGGATCGTGCGTGCTCGTAGGTGCTGCCGCCTTGTGGCGGTTCTGGCG
MAACWDQTRRCCKELPPCGGSG	115	ATGGCAGCGTGTTGGGATTAGACTCGTAGGTGCTGTAAGGAGTTGCCTCCGTGTGGCGGTTCTGGCG
MAACSSRKVCCLASDVTCGGSG	113	ATGGCAGCGTGTAGTAGTAGGAAGGTGTGTTGCCTGGCTAGTGATGTTACTTGTGGCGGTTCTGGCG
MAACPMVLPPCQHTLHECGGSG	111	ATGGCAGCGTGTCCTATGGTTCTGCCGCCTTGCTAGCATACTCTGCATGAGTGTGGCGGTTCTGGCG
MAACRGILPPCAPQAYECGGSG	109	ATGGCAGCGTGTCGTGGGATTTTGCCTCCGTGCGCTCCTTAGGCGTATGAGTGTGGCGGTTCTGGCG
MAACSSQLPPCDRVQELGGSG	108	ATGGCAGCGTGTTCGTCTCAGCTGCCGCCGTGCGATCGCGTGTAGGAATTGGGCGGTTCTGGCG
MAACRYTQESCPHILPPCGGSG	107	ATGGCAGCGTGTCGTTATACGTAGGAGAGTTGCCCTCATATTTTGCCTCCGTGTGGCGGTTCTGGCG
MAACWDKRVCCVAPWRPCGGSG	104	ATGGCAGCGTGTTGGGATAAGCGTGTGTGTTGCGTGGCGCCTTGGCGTCCTTGTGGCGGTTCTGGCG
MAACQGTLPPCPAGIKPCGGSG	101	ATGGCAGCGTGTTAGGGGACGCTTCCGCCTTGCCCGGCGGGTATTAAGCCTTGTGGCGGTTCTGGCG
MAACDTYQEECTHILPPCGGSG	98	ATGGCAGCGTGTGATACGTATTAGGAGGAGTGCACTCATATTTTGCCGCCTTGTGGCGGTTCTGGCG
MAACPNLLPPCGGSQTACGGSG	97	ATGGCAGCGTGTCCTAATTTGCTTCCGCCTTGCGGGGGGTCTTAGACGGCTTGTGGCGGTTCTGGCG
MAACSYLEQTCNQYLPPCGGSG	91	ATGGCAGCGTGTAGTTATCTTGAGCAGACGTGCAATTAGTATCTTCCTCCTTGTGGCGGTTCTGGCG
MAACTSVLPPCSGNEQLCGGSG	90	ATGGCAGCGTGTACTTCGGTGCTTCCTCCGTGCTCTGGTAATGAGTAGCTTTGTGGCGGTTCTGGCG
MAACSLLPPQCTFGSPSCGGSG	90	ATGGCAGCGTGTAGTCTGTTGCCTCCGTAGTGCACGTTTGGTTCGCCGTCTTGTGGCGGTTCTGGCG
MAACTQHGRQCGHFLPPCGGSG	88	ATGGCAGCGTGTACGTAGCATGGGCGGCAGTGCGGGCATTTTCTGCCTCCGTGTGGCGGTTCTGGCG
MAACLRVLGCCVSVERPCGGSG	88	ATGGCAGCGTGTTTGAGGGTTCTTGGGTGTTGCGTGAGTGTTGAGCGTCCGTGTGGCGGTTCTGGCG
MAACVLFLPPCPQGSHLCGGSG	87	ATGGCAGCGTGTGTGCTTTTTCTTCCGCCGTGCCCTTAGGGGTCTCATCTGTGTGGCGGTTCTGGCG
MAACGRYLPPCPQSPHNCGGSG	86	ATGGCAGCGTGTGGTAGGTATCTTCCGCCGTGCCCTCAGAGTCCTCATAATTGTGGCGGTTCTGGCG
MAACGSARKCCLSVLPPCGGSG	86	ATGGCAGCGTGTGGGAGTGCGCGTAAGTGTTGCTTGTCTGTTCTGCCTCCTTGTGGCGGTTCTGGCG
MAACLPLLPPCDGDCLQCGGSG	83	ATGGCAGCGTGTTTGCCGTTGCTGCCGCCTTGCGATGGTGATTGTCTTCAGTGTGGCGGTTCTGGCG
MAACMTGQGRCGGVLPPCGGSG	82	ATGGCAGCGTGTATGACGGGTTAGGGGCGTTGCGGTGGGGTGTTGCCTCCTTGTGGCGGTTCTGGCG
MAACHGRQVCCLAPARPCGGSG	81	ATGGCAGCGTGTCATGGTCGGCAGGTGTGTTGCCTTGCGCCTGCTCGGCCGTGTGGCGGTTCTGGCG
MAACLLGPRECGQLPPLCGGSG	81	ATGGCAGCGTGTTTGCTGGGTCCGAGGGAGTGCGGTTAGTTGCCGCCTCTGTGTGGCGGTTCTGGCG
MAACEGVLPPCPQLTQYCGGSG	81	ATGGCAGCGTGTGAGGGGGTGTTGCCTCCTTGCCCGTAGCTGACGCAGTATTGTGGCGGTTCTGGCG
MAACTWINGVCDGILPPCGGSG	80	ATGGCAGCGTGTACTTGGATTAATGGTGTGTGCGATGGTATTCTGCCGCCTTGTGGCGGTTCTGGCG
MAACPGLLPPCQLGRPGCGGSG	79	ATGGCAGCGTGTCCGGGTCTGTTGCCTCCGTGCTAGCTTGGTCGTCCGGGTTGTGGCGGTTCTGGCG
MAACRQWLPPCPVGGTACGGSG	79	ATGGCAGCGTGTAGGCAGTGGCTGCCTCCTTGCCCTGTTGGGGGGACGGCTTGTGGCGGTTCTGGCG
MAACGIAGLPCSQVLPPCGGSG	77	ATGGCAGCGTGTGGGATTGCTGGTCTTCCTTGCTCTTAGGTGCTTCCTCCTTGTGGCGGTTCTGGCG
MAACTCYGKKCLNVLPPCGGSG	76	ATGGCAGCGTGTACTTGTTATGGTAAGAAGTGCCTTAATGTTCTGCCTCCTTGTGGCGGTTCTGGCG
MAACGQILPPCNFQINGCGGSG	76	ATGGCAGCGTGTGGTTAGATTCTTCCTCCTTGCAATTTTTAGATTAATGGTTGTGGCGGTTCTGGCG
MAACKHLLPPCDYMSQNCGGSG	73	ATGGCAGCGTGTAAGCATCTGCTTCCTCCGTGCGATTATATGAGTTAGAATTGTGGCGGTTCTGGCG
MAACHSDVQPCPNILPPCGGSG	73	ATGGCAGCGTGTCATAGTGATGTTTAGCCGTGCCCGAATATTCTGCCTCCTTGTGGCGGTTCTGGCG
MAACEYELQLCGNTLPPCGGSG	72	ATGGCAGCGTGTGAGTATGAGTTGTAGCTTTGCGGTAATACTCTTCCGCCGTGTGGCGGTTCTGGCG
MAACVGRLPNCGLPPWQCGGSG	72	ATGGCAGCGTGTGTTGGTCGTCTGCCTAATTGCGGTCTTCCTCCTTGGTAGTGTGGCGGTTCTGGCG
MAACVSRRVCCLDIAHQCGGSG	69	ATGGCAGCGTGTGTTTCGAGGCGGGTTTGTTGCCTGGATATTGCTCATCAGTGTGGCGGTTCTGGCG
MAACQARFDRCGLPPWNCGGSG	69	ATGGCAGCGTGTTAGGCTAGGTTTGATCGTTGCGGTCTTCCGCCGTGGAATTGTGGCGGTTCTGGCG
MAACTQYPQSCAGILPPCGGSG	68	ATGGCAGCGTGTACGTAGTATCCGCAGTCGTGCGCTGGGATTCTTCCGCCGTGTGGCGGTTCTGGCG
MAACGIQRVCCLAPDVACGGSG	67	ATGGCAGCGTGTGGTATTTAGAGGGTGTGTTGCTTGGCTCCTGATGTTGCTTGTGGCGGTTCTGGCG
MAACNSAVQKCCGQLPPCGGSG	67	ATGGCAGCGTGTAATTCTGCTGTGCAGAAGTGCTGTGGGTAGCTTCCGCCTTGTGGCGGTTCTGGCG
MAACWHKQMCCVDIKAPCGGSG	67	ATGGCAGCGTGTTGGCATAAGTAGATGTGTTGCGTTGATATTAAGGCTCCGTGTGGCGGTTCTGGCG

MAACGGLLPPCGLSLQECGGSG	66	ATGGCAGCGTGTGGGGGGTTGCTGCCGCCGTGCGGTTTGAGTCTGTAGGAGTGTGGCGGTTCTGGCG
MAACWWQAKRCGLILPPCGGSG	65	ATGGCAGCGTGTTGGTGGCAGGCGAAGCGGTGCGGGTTGATTCTTCCTCCTTGTGGCGGTTCTGGCG
MAACFPKAVCCLASELLCGGSG	65	ATGGCAGCGTGTTTTCCGAAGGCTGTGTGTTGCTTGGCTTCTGAGTTGTTGTGTGGCGGTTCTGGCG
MAACPLILPPCGLSGRNCGGSG	65	ATGGCAGCGTGTCCGCTTATTTTGCCTCCGTGCGGTCTGTCTGGGCGTAATTGTGGCGGTTCTGGCG
MAACNVWEQHCQNILPPCGGSG	64	ATGGCAGCGTGTAATGTTTGGGAGCAGCATTGCTAGAATATTCTTCCTCCTTGTGGCGGTTCTGGCG
MAACGGQLPPCGVVYSSCGGSG	63	ATGGCAGCGTGTGGTGGGTAGCTTCCGCCGTGCGGGGTGGTTTATTCTAGTTGTGGCGGTTCTGGCG
MAACDSQVKKCANLLPPCGGSG	63	ATGGCAGCCTGTGATTCTTAGGTGAAGAAGTGCGCTAATCTGCTGCCTCCTTGTGGCGGTTCTGGCG
MAACEYRVDPCGQLLPPCGGSG	63	ATGGCAGCGTGTGAGTATCGTGTGGATCCGTGCGGGTAGTTGTTGCCTCCGTGTGGCGGTTCTGGCG
MAACGVVLPPCPQGMNWCGGSG	63	ATGGCAGCGTGTGGTGTGGTGCTGCCGCCTTGCCCTTAGGGTATGAATTGGTGTGGCGGTTCTGGCG
MAACTGVLPPCSYKSERCGGSG	63	ATGGCAGCGTGTACTGGTGTGTTGCCGCCGTGCTCTTATAAGAGTGAGCGTTGTGGCGGTTCTGGCG
MAACSGVLPPCSGRMQSCGGSG	63	ATGGCAGCGTGTAGTGGGGTGCTGCCTCCTTGCAGTGGGAGGATGTAGTCGTGTGGCGGTTCTGGCG
MAACRGVLPPCNSAQVGCGGSG	63	ATGGCAGCGTGTAGGGGGGTGCTGCCTCCGTGCAATAGTGCGTAGGTGGGGTGTGGCGGTTCTGGCG
MAACTRPQDACPHILPPCGGSG	62	ATGGCAGCGTGTACGCGGCCGTAGGATGCGTGCCCTCATATTCTGCCTCCGTGTGGCGGTTCTGGCG
MAACPRILPPCASQAPLCGGSG	62	ATGGCAGCGTGTCCTCGGATTCTGCCGCCTTGCGCGAGTTAGGCTCCGCTTTGTGGCGGTTCTGGCG

References

1 Craik, D. J., Fairlie, D. P., Liras, S. & Price, D. The future of peptide-based drugs. *Chem Biol Drug Des* **81**, 136-147, (2013).

2 Latham, P. W. Therapeutic peptides revisited. *Nat Biotechnol* **17**, 755-757, (1999).

3 Hamzeh-Mivehroud, M., Alizadeh, A. A., Morris, M. B., Church, W. B. & Dastmalchi, S. Phage display as a technology delivering on the promise of peptide drug discovery. *Drug Discov Today* **18**, 1144-1157, (2013).

4 Ladner, R. C., Sato, A. K., Gorzelany, J. & de Souza, M. Phage display-derived peptides as therapeutic alternatives to antibodies. *Drug Discov Today* **9**, 525-529, (2004).

5 Macdougall, I. C., Provenzano, R., Sharma, A., Spinowitz, B. S., Schmidt, R. J., Pergola, P. E., Zabaneh, R. I., Tong-Starksen, S., Mayo, M. R., Tang, H., Polu, K. R., Duliege, A. M. & Fishbane, S. Peginesatide for anemia in patients with chronic kidney disease not receiving dialysis. *N Engl J Med* **368**, 320-332, (2013).

6 Fishbane, S., Schiller, B., Locatelli, F., Covic, A. C., Provenzano, R., Wiecek, A., Levin, N. W., Kaplan, M., Macdougall, I. C., Francisco, C., Mayo, M. R., Polu, K. R., Duliege, A. M. & Besarab, A. Peginesatide in patients with anemia undergoing hemodialysis. *N Engl J Med* **368**, 307-319, (2013).

7 Bennett, C. L., Jacob, S., Hymes, J., Usvyat, L. A. & Maddux, F. W. Anaphylaxis and hypotension after administration of peginesatide. *N Engl J Med* **370**, 2055-2056, (2014).

8 Yamagishi, Y., Shoji, I., Miyagawa, S., Kawakami, T., Katoh, T., Goto, Y. & Suga, H. Natural product-like macrocyclic N-methyl-peptide inhibitors against a ubiquitin ligase uncovered from a ribosome-expressed de novo library. *Chem Biol* **18**, 1562-1570, (2011).

9 Giordanetto, F. & Kihlberg, J. Macrocyclic drugs and clinical candidates: what can medicinal chemists learn from their properties? *J Med Chem* **57**, 278-295, (2014).

10 Driggers, E. M., Hale, S. P., Lee, J. & Terrett, N. K. The exploration of macrocycles for drug discovery--an underexploited structural class. *Nat Rev Drug Discov* **7**, 608-624, (2008).

11 Gongora-Benitez, M., Tulla-Puche, J. & Albericio, F. Multifaceted roles of disulfide bonds. Peptides as therapeutics. *Chem Rev* **114**, 901-926, (2014).

12 Bock, J. E., Gavenonis, J. & Kritzer, J. A. Getting in shape: controlling peptide bioactivity and bioavailability using conformational constraints. *ACS Chem Biol* **8**, 488-499, (2013).

13 Smith, G. P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315-1317, (1985).

14 McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**, 552-554, (1990).

15 Smith, G. P. & Petrenko, V. a. Phage Display. *Chem Rev* **97**, 391-410, (1997).

16 Kay, B. K., Winter, J. & McCafferty, J. *Phage Display of Peptides and Proteins: A Laboratory Manual*. (Academic Press, 1996).

17 Barbas, C. F. I., Burton, D., Scott, J. & Silverman, G. *Phage Display : A Laboratory Manual*. (Cold Spring Harbor Laboratory Press, 2001).

18 Cabilly, S. The basic structure of filamentous phage and its use in the display of combinatorial peptide libraries. *Mol Biotechnol* **12**, 143-148, (1999).

19 Burritt, J. B., Bond, C. W., Doss, K. W. & Jesaitis, a. J. Filamentous phage display of oligopeptide libraries. *Anal Biochem* **238**, 1-13, (1996).

20 Clackson, T. & Lowman, H. *Phage Display: A Practical Approach (The Practical Approach Series, 266)*. (Oxford University Press, UK, 2004).

21 Parmley, S. F. & Smith, G. P. Antibody-selectable filamentous fd phage vectors: affinity purification of target genes. *Gene* **73**, 305-318, (1988).

22 Bain, B. & Brazil, M. Adalimumab. *Nat Rev Drug Discov* **2**, 693-694, (2003).

23 Sanz, I., Yasothan, U. & Kirkpatrick, P. Belimumab. *Nat Rev Drug Discov* **10**, 335-336, (2011).

24 Fox, J. L. Anthrax drug first antibacterial mAb to win approval. *Nat Biotechnol* **31**, 8, (2013).

25 Rowley, M. J., O'Connor, K. & Wijeyewickrema, L. Phage display for epitope determination: a paradigm for identifying receptor-ligand interactions. *Biotechnol Annu Rev* **10**, 151-188, (2004).

26 McLafferty, M. A., Kent, R. B., Ladner, R. C. & Markland, W. M13 bacteriophage displaying disulfide-constrained microproteins. *Gene* **128**, 29-36, (1993).

27 O'Neil, K. T., Hoess, R. H., Jackson, S. A., Ramachandran, N. S., Mousa, S. A. & DeGrado, W. F. Identification of novel peptide antagonists for GPIIb/IIIa from a conformationally constrained phage peptide library. *Proteins* **14**, 509-515, (1992).

28 Lowman, H. B. Bacteriophage display and discovery of peptide leads for drug development. *Annu Rev Biophys Biomol Struct* **26**, 401-424, (1997).

29 Nord, K., Nilsson, J., Nilsson, B., Uhlen, M. & Nygren, P. A. A combinatorial library of an alpha-helical bacterial receptor domain. *Protein Eng* **8**, 601-608, (1995).

30 Koide, A., Bailey, C. W., Huang, X. & Koide, S. The fibronectin type III domain as a scaffold for novel binding proteins. *J Mol Biol* **284**, 1141-1151, (1998).

31 Beste, G., Schmidt, F. S., Stibora, T. & Skerra, A. Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold. *Proc Natl Acad Sci USA* **96**, 1898-1903, (1999).

32 Amstutz, P., Binz, H. K., Parizek, P., Stumpp, M. T., Kohl, A., Grutter, M. G., Forrer, P. & Pluckthun, A. Intracellular kinase inhibitors selected from combinatorial libraries of designed ankyrin repeat proteins. *J Biol Chem* **280**, 24715-24722, (2005).

33 Binz, H. K., Amstutz, P. & Pluckthun, A. Engineering novel binding proteins from nonimmunoglobulin domains. *Nat Biotechnol* **23**, 1257-1268, (2005).

34 Getmanova, E. V., Chen, Y., Bloom, L., Gokemeijer, J., Shamah, S., Warikoo, V., Wang, J., Ling, V. & Sun, L. Antagonists to human and mouse vascular endothelial growth factor receptor 2 generated by directed protein evolution in vitro. *Chem Biol* **13**, 549-556, (2006).

35 Souied, E. H., Devin, F., Mauget-Faysse, M., Kolar, P., Wolf-Schnurrbusch, U., Framme, C., Gaucher, D., Querques, G., Stumpp, M. T. & Wolf, S. Treatment of exudative age-related macular degeneration with a designed ankyrin repeat protein that binds vascular endothelial growth factor: a Phase I/II study. *Am J Ophthalmol*, (2014).

36 Ng, S., Jafari, M. R. & Derda, R. Bacteriophages and viruses as a support for organic synthesis and combinatorial chemistry. *ACS Chem Biol* **7**, 123-138, (2012).

37 Angelini, A. & Heinis, C. Post-translational modification of genetically encoded polypeptide libraries. *Curr Opin Chem Biol* **15**, 355-361, (2011).

38 Kather, I., Bippes, C. A. & Schmid, F. X. A stable disulfide-free gene-3-protein of phage fd generated by in vitro evolution. *J Mol Biol* **354**, 666-678, (2005).

39 Arai, K., Tsutsumi, H. & Mihara, H. A monosaccharide-modified peptide phage library for screening of ligands to carbohydrate-binding proteins. *Bioorg Med Chem Lett* **23**, 4940-4943, (2013).

40 Jespers, L., Bonnert, T. P. & Winter, G. Selection of optical biosensors from chemisynthetic antibody libraries. *Protein Eng Des Sel* **17**, 709-713, (2004).

41 Jafari, M. R., Deng, L., Kitov, P. I., Ng, S., Matochko, W. L., Tjhung, K. F., Zeberoff, A., Elias, A., Klassen, J. S. & Derda, R. Discovery of light-responsive ligands through screening of a light-responsive genetically encoded library. *ACS Chem Biol* **9**, 443-450, (2014).

42 Bellotto, S., Chen, S., Rentero Rebollo, I., Wegner, H. A. & Heinis, C. Phage selection of photoswitchable peptide ligands. *J Am Chem Soc* **136**, 5880-5883, (2014).

43 Heinis, C., Rutherford, T., Freund, S. & Winter, G. Phage-encoded combinatorial chemical libraries based on bicyclic peptides. *Nat Chem Biol* **5**, 502-507, (2009).

44 Angelini, A., Cendron, L., Chen, S., Touati, J., Winter, G., Zanotti, G. & Heinis, C. Bicyclic peptide inhibitor reveals large contact interface with a protease target. *ACS Chem Biol* **7**, 817-821, (2012).

45 Ng, S., Jafari, M. R., Matochko, W. L. & Derda, R. Quantitative synthesis of genetically encoded glycopeptide libraries displayed on M13 phage. *ACS Chem Biol* **7**, 1482-1487, (2012).

46 Hess, G. T., Cragnolini, J. J., Popp, M. W., Allen, M. A., Dougan, S. K., Spooner, E., Ploegh, H. L., Belcher, A. M. & Guimaraes, C. P. M13 bacteriophage display framework that allows sortase-mediated modification of surface-accessible phage proteins. *Bioconj Chem* **23**, 1478-1487, (2012).

47 Roberts, R. W. & Szostak, J. W. RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc Natl Acad Sci USA* **94**, 12297-12302, (1997).

48 Hanes, J. & Pluckthun, A. In vitro selection and evolution of functional proteins by using ribosome display. *Proc Natl Acad Sci USA* **94**, 4937-4942, (1997).

49 Liu, C. C., Mack, A. V., Tsao, M. L., Mills, J. H., Lee, H. S., Choe, H., Farzan, M., Schultz, P. G. & Smider, V. V. Protein evolution with an expanded genetic code. *Proc Natl Acad Sci USA* **105**, 17688-17693, (2008).

50 Goto, Y., Katoh, T. & Suga, H. Flexizymes for genetic code reprogramming. *Nat Protoc* **6**, 779-790, (2011).

51 Murakami, H., Ohta, A., Goto, Y., Sako, Y. & Suga, H. Flexizyme as a versatile tRNA acylation catalyst and the application for translation. *Nucleic Acids Symp Ser (Oxf)*, 35-36, (2006).

52 Ohuchi, M., Murakami, H. & Suga, H. The flexizyme system: a highly flexible tRNA aminoacylation tool for the translation apparatus. *Curr Opin Chem Biol* **11**, 537-542, (2007).

53 Mylne, J. S., Colgrave, M. L., Daly, N. L., Chanson, A. H., Elliott, A. G., McCallum, E. J., Jones, A. & Craik, D. J. Albumins and their processing machinery are hijacked for cyclic peptides in sunflower. *Nat Chem Biol* **7**, 257-259, (2011).

54 Craik, D. J., Simonsen, S. & Daly, N. L. The cyclotides: novel macrocyclic peptides as scaffolds in drug design. *Curr Opin Drug Discov Devel* **5**, 251-260, (2002).

55 Ganz, T. Defensins: antimicrobial peptides of innate immunity. *Nat Rev Immunol* **3**, 710-720, (2003).

56 Timmerman, P., Beld, J., Puijk, W. C. & Meloen, R. H. Rapid and quantitative cyclization of multiple peptide loops onto synthetic scaffolds for structural mimicry of protein surfaces. *ChemBioChem* **6**, 821-824, (2005).

57 Timmerman, P., Puijk, W. C., Boshuizen, R. S., Dijken, P. V., Slootstra, J. W., Beurskens, F. J., Parren, P. W. H. I., Huber, A., Bachmann, M. F. & Meloen, R. H. Functional reconstruction of structurally complex epitopes using CLIPS technology. *The Open Vaccine Journal* **2**, 56-67, (2009).

58 Baeriswyl, V., Calzavarini, S., Gerschheimer, C., Diderich, P., Angelillo-Scherrer, A. & Heinis, C. Development of a selective peptide macrocycle inhibitor of coagulation factor XII toward the generation of a safe antithrombotic therapy. *J Med Chem* **56**, 3742-3746, (2013).

59 Diderich, P. & Heinis, C. Phage selection of bicyclic peptides binding Her2. *Tetrahedron*, In Press, (2014).

60 Chen, S., Morales-Sanfrutos, J., Angelini, A., Cutting, B. & Heinis, C. Structurally diverse cyclisation linkers impose different backbone conformations in bicyclic peptides. *ChemBioChem* **13**, 1032-1038, (2012).

61 Chen, S., Bertoldo, D., Angelini, A., Pojer, F. & Heinis, C. Peptide ligands stabilized by small molecules. *Angew Chem Int Ed Engl* **53**, 1602-1606, (2014).

62 Baeriswyl, V., Rapley, H., Pollaro, L., Stace, C., Teufel, D., Walker, E., Chen, S. Y., Winter, G., Tite, J. & Heinis, C. Bicyclic Peptides with Optimized Ring Size Inhibit Human Plasma Kallikrein and its Orthologues While Sparing Paralogous Proteases. *ChemMedChem* **7**, 1173-1176, (2012).

63 Chen, S., Rentero Rebollo, I., Buth, S. A., Morales-Sanfrutos, J., Touati, J., Leiman, P. G. & Heinis, C. Bicyclic peptide ligands pulled out of cysteine-rich peptide libraries. *J Am Chem Soc* **135**, 6562-6569, (2013).

64 Andreasen, P. A., Egelund, R. & Petersen, H. H. The plasminogen activation system in tumor growth, invasion, and metastasis. *Cell Mol Life Sci* **57**, 25-40, (2000).

65 Chen, S., Gfeller, D., Buth, S. A., Michielin, O., Leiman, P. G. & Heinis, C. Improving binding affinity and stability of peptide ligands by substituting glycines with D-amino acids. *ChemBioChem* **14**, 1316-1322, (2013).

66 Angelini, A., Morales-Sanfrutos, J., Diderich, P., Chen, S. Y. & Heinis, C. Bicyclization and tethering to albumin yields long-acting peptide antagonists. *J Med Chem* **55**, 10187-10197, (2012).

67 Angelini, A., Diderich, P., Morales-Sanfrutos, J., Thurnheer, S., Hacker, D., Menin, L. & Heinis, C. Chemical macrocyclization of peptides fused to antibody Fc fragments. *Bioconj Chem* **23**, 1856-1863, (2012).

68 Dennis, M. S., Zhang, M., Meng, Y. G., Kadkhodayan, M., Kirchhofer, D., Combs, D. & Damico, L. a. Albumin binding as a general strategy for improving the pharmacokinetics of proteins. *J Biol Chem* **277**, 35035-35043, (2002).

69 Pollaro, L. *Therapeutic effect and biodistribution of a bicyclic peptide inhibitor of urokinase-type plasminogen activator*, École polytechnique fédérale de Lausanne EPFL, n° 6181, (2014).

70 Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, C. A., Hutchison, C. A., Slocombe, P. M. & Smith, M. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* **265**, 687-695, (1977).

71 Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. & Hood, L. E. Fluorescence detection in automated DNA sequence analysis. *Nature* **321**, 674-679, (1986).

72 Mardis, E. R. Next-generation sequencing platforms. *Annu Rev Anal Chem* **6**, 287-303, (2013).

73 Schuster, S. C. Next-generation sequencing transforms today's biology. *Nat Methods* **5**, 16-18, (2008).

74 Mannocci, L., Zhang, Y., Scheuermann, J., Leimbacher, M., De Bellis, G., Rizzi, E., Dumelin, C., Melkko, S. & Neri, D. High-throughput sequencing allows the identification of binding molecules isolated from DNA-encoded chemical libraries. *Proc Natl Acad Sci U S A* **105**, 17670-17675, (2008).

75 Matochko, W. L., Chu, K., Jin, B., Lee, S. W., Whitesides, G. M. & Derda, R. Deep sequencing analysis of phage libraries using Illumina platform. *Methods* **58**, 47-55, (2012).

76 Ravn, U., Gueneau, F., Baerlocher, L., Osteras, M., Desmurs, M., Malinge, P., Magistrelli, G., Farinelli, L., Kosco-Vilbois, M. H. & Fischer, N. By-passing in vitro screening--next generation sequencing technologies applied to antibody display and in silico candidate selection. *Nucleic Acids Res* **38**, e193, (2010).

77 Olson, C. A., Nie, J., Diep, J., Al-Shyoukh, I., Takahashi, T. T., Al-Mawsawi, L. Q., Bolin, J. M., Elwell, A. L., Swanson, S., Stewart, R., Thomson, J. A., Soh, H. T., Roberts, R. W. & Sun, R. Single-round, multiplexed antibody mimetic design through mRNA display. *Angew Chem Int Ed Engl* **51**, 12449-12453, (2012).

78 Larman, H. B., Xu, G. J., Pavlova, N. N. & Elledge, S. J. Construction of a rationally designed antibody platform for sequencing-assisted selection. *Proc Natl Acad Sci USA* **109**, 18523-18528, (2012).

79 Matochko, W. L., Li, S. C., Tang, S. K. Y. & Derda, R. Prospective identification of parasitic sequences in phage display screens. *Nucleic Acids Res* **42**, 1784-1798, (2014).

80 Kawakami, T., Ishizawa, T. & Murakami, H. Extensive reprogramming of the genetic code for genetically encoded synthesis of highly N-alkylated polycyclic peptidomimetics. *J Am Chem Soc* **135**, 12297-12304, (2013).

81 Naso, M. F., Lu, J. & Panavas, T. Deep sequencing approaches to antibody discovery. *Curr Drug Discov Technol* **11**, 85-95, (2014).

82 t Hoen, P. A., Jirka, S. M., Ten Broeke, B. R., Schultes, E. A., Aguilera, B., Pang, K. H., Heemskerk, H., Aartsma-Rus, A., van Ommen, G. J. & den Dunnen, J. T. Phage display screening without repetitious selection rounds. *Anal Biochem* **421**, 622-631, (2012).

83 WHO. Antimicrobial resistance: global report on surveillance 2014. (2014).

84 Payne, D. J., Gwynn, M. N., Holmes, D. J. & Pompliano, D. L. Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat Rev Drug Discov* **6**, 29-40, (2007).

85 Huang, J. X., Bishop-Hurley, S. L. & Cooper, M. A. Development of anti-Infectives using phage display: biological agents against bacteria, viruses, and parasites. *Antimicrobial agents and chemotherapy* **56**, 4569-4582, (2012).

86 Hancock, R. E. & Sahl, H. G. Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nat Biotechnol* **24**, 1551-1557, (2006).

87 Melo, M. N., Ferre, R. & Castanho, M. A. Antimicrobial peptides: linking partition, activity and high membrane-bound concentrations. *Nat Rev Microbiol* **7**, 245-250, (2009).

88 Wiedemann, I., Breukink, E., van Kraaij, C., Kuipers, O. P., Bierbaum, G., de Kruijff, B. & Sahl, H. G. Specific binding of nisin to the peptidoglycan precursor lipid II combines pore formation and inhibition of cell wall biosynthesis for potent antibiotic activity. *J Biol Chem* **276**, 1772-1779, (2001).

89 Wilson, D. N. Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat Rev Microbiol* **12**, 35-48, (2014).

90 Paradis-Bleau, C., Lloyd, A., Sanschagrin, F., Clarke, T., Blewett, A., Bugg, T. D. H. & Levesque, R. C. Phage display-derived inhibitor of the essential cell wall biosynthesis enzyme MurF. *BMC Biochemistry* **9**, 33, (2008).

91 Houimel, M., Mach, J. P., Corthesy-Theulaz, I., Corthesy, B. & Fisch, I. New inhibitors of Helicobacter pylori urease holoenzyme selected from phage-displayed peptide libraries. *Eur J Biochem* **262**, 774-780, (1999).

92 Srinivas, N., Jetter, P., Ueberbacher, B. J., Werneburg, M., Zerbe, K., Steinmann, J., Van der Meijden, B., Bernardini, F., Lederer, A., Dias, R. L., Misson, P. E., Henze, H., Zumbrunn, J., Gombert, F. O., Obrecht, D., Hunziker, P., Schauer, S., Ziegler, U., Kach, A., Eberl, L., Riedel, K., DeMarco, S. J. & Robinson, J. A. Peptidomimetic antibiotics target outer-membrane biogenesis in Pseudomonas aeruginosa. *Science* **327**, 1010-1013, (2010).

93 Borg, M. A., Hulscher, M., Scicluna, E. A., Richards, J., Azanowsky, J. M., Xuereb, D., Huis, A., Moro, M. L., Maltezou, H. C. & Frank, U. Prevention of meticillin-resistant Staphylococcus aureus bloodstream infections in European hospitals: moving beyond policies. *J Hosp Infect* **87**, 203-211, (2014).

94 Fridkin, S. K., Hageman, J. C., Morrison, M., Sanza, L. T., Como-Sabetti, K., Jernigan, J. A., Harriman, K., Harrison, L. H., Lynfield, R. & Farley, M. M. Methicillin-resistant staphylococcus aureus disease in three communities. *N Engl J Med* **352**, 1436-1444, (2005).

95 Weigel, L. M., Clewell, D. B., Gill, S. R., Clark, N. C., McDougal, L. K., Flannagan, S. E., Kolonay, J. F., Shetty, J., Killgore, G. E. & Tenover, F. C. Genetic analysis of a high-level vancomycin-resistant isolate of Staphylococcus aureus. *Science* **302**, 1569-1571, (2003).

96 Salgado-Pabon, W. & Schlievert, P. M. Models matter: the search for an effective Staphylococcus aureus vaccine. *Nat Rev Microbiol* **12**, 585-591, (2014).

97 Oliveira, D. C., Tomasz, A. & de Lencastre, H. Secrets of success of a human pathogen: molecular evolution of pandemic clones of meticillin-resistant Staphylococcus aureus. *Lancet Infect Dis* **2**, 180-189, (2002).

98 Clatworthy, A. E., Pierson, E. & Hung, D. T. Targeting virulence: a new paradigm for antimicrobial therapy. *Nat Chem Biol* **3**, 541-548, (2007).

99 Cegelski, L., Marshall, G. R., Eldridge, G. R. & Hultgren, S. J. The biology and future prospects of antivirulence therapies. *Nat Rev Microbiol* **6**, 17-27, (2008).

100 Mazmanian, S. K., Liu, G., Ton-That, H. & Schneewind, O. Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall. *Science* **285**, 760-763, (1999).

101 Mazmanian, S. K., Ton-That, H., Su, K. & Schneewind, O. An iron-regulated sortase anchors a class of surface protein during Staphylococcus aureus pathogenesis. *Proc Natl Acad Sci USA* **99**, 2293-2298, (2002).

102 Mazmanian, S. K., Liu, G., Jensen, E. R., Lenoy, E. & Schneewind, O. Staphylococcus aureus sortase mutants defective in the display of surface proteins and in the pathogenesis of animal infections. *Proc Natl Acad Sci USA* **97**, 5510-5515, (2000).

103 Jonsson, I. M., Mazmanian, S. K., Schneewind, O., Bremell, T. & Tarkowski, A. The role of Staphylococcus aureus sortase A and sortase B in murine arthritis. *Microbes Infect* **5**, 775-780, (2003).

104 Jonsson, I. M., Mazmanian, S. K., Schneewind, O., Verdrengh, M., Bremell, T. & Tarkowski, A. On the role of Staphylococcus aureus sortase and sortase-catalyzed surface protein anchoring in murine septic arthritis. *J Infect Dis* **185**, 1417-1424, (2002).

105 Weiss, W. J., Lenoy, E., Murphy, T., Tardio, L., Burgio, P., Projan, S. J., Schneewind, O. & Alksne, L. Effect of srtA and srtB gene expression on the virulence of Staphylococcus aureus in animal models of infection. *J Antimicrob Chemother* **53**, 480-486, (2004).

106 Wardenburg, J. B., Patel, R. J. & Schneewind, O. Surface proteins and exotoxins are required for the pathogenesis of Staphylococcus aureus pneumonia. *Infect Immun* **75**, 1040-1044, (2007).

107 Ton-That, H., Mazmanian, S. K., Alksne, L. & Schneewind, O. Anchoring of surface proteins to the cell wall of Staphylococcus aureus. Cysteine 184 and histidine 120 of sortase form a thiolate-imidazolium ion pair for catalysis. *J Biol Chem* **277**, 7447-7452, (2002).

108 Frankel, B. A., Kruger, R. G., Robinson, D. E., Kelleher, N. L. & McCafferty, D. G. Staphylococcus aureus sortase transpeptidase SrtA: insight into the kinetic mechanism and evidence for a reverse protonation catalytic mechanism. *Biochemistry* **44**, 11188-11200, (2005).

109 Clancy, K. W., Melvin, J. A. & McCafferty, D. G. Sortase transpeptidases: insights into mechanism, substrate specificity, and inhibition. *Biopolymers* **94**, 385-396, (2010).

110 Bronner, S., Monteil, H. & Prevost, G. Regulation of virulence determinants in Staphylococcus aureus: complexity and applications. *FEMS Microbiol Rev* **28**, 183-200, (2004).

111 Suree, N., Yi, S. W., Thieu, W., Marohn, M., Damoiseaux, R., Chan, A., Jung, M. E. & Clubb, R. T. Discovery and structure-activity relationship analysis of Staphylococcus aureus sortase A inhibitors. *Bioorg Med Chem* **17**, 7174-7185, (2009).

112 Jang, K. H., Chung, S. C., Shin, J., Lee, S. H., Kim, T. I., Lee, H. S. & Oh, K. B. Aaptamines as sortase A inhibitors from the tropical sponge Aaptos aaptos. *Bioorg Med Chem Lett* **17**, 5366-5369, (2007).

113 Oh, K. B., Mar, W., Kim, S., Kim, J. Y., Oh, M. N., Kim, J. G., Shin, D., Sim, C. J. & Shin, J. Bis(indole) alkaloids as sortase A inhibitors from the sponge Spongosorites sp. *Bioorg Med Chem Lett* **15**, 4927-4931, (2005).

114 Park, B. S., Kim, J. G., Kim, M. R., Lee, S. E., Takeoka, G. R., Oh, K. B. & Kim, J. H. Curcuma longa L. constituents inhibit sortase A and Staphylococcus aureus cell adhesion to fibronectin. *J Agric Food Chem* **53**, 9005-9009, (2005).

115 Tsukamoto, S., Yamanokuchi, R., Yoshitomi, M., Sato, K., Ikeda, T., Rotinsulu, H., Mangindaan, R. E., de Voogd, N. J., van Soest, R. W. & Yokosawa, H. Aaptamine, an alkaloid from the sponge Aaptos suberitoides, functions as a proteasome inhibitor. *Bioorg Med Chem Lett* **20**, 3341-3343, (2010).

116 Oh, K. B., Mar, W., Kim, S., Kim, J. Y., Lee, T. H., Kim, J. G., Shin, D., Sim, C. J. & Shin, J. Antimicrobial activity and cytotoxicity of bis(indole) alkaloids from the sponge Spongosorites sp. *Biol Pharm Bull* **29**, 570-573, (2006).

117 Duvoix, A., Blasius, R., Delhalle, S., Schnekenburger, M., Morceau, F., Henry, E., Dicato, M. & Diederich, M. Chemopreventive and therapeutic effects of curcumin. *Cancer Lett* **223**, 181-190, (2005).

118 Oh, K. B., Kim, S. H., Lee, J., Cho, W. J., Lee, T. & Kim, S. Discovery of diarylacrylonitriles as a novel series of small molecule sortase A inhibitors. *J Med Chem* **47**, 2418-2421, (2004).

119 Oh, K. B., Nam, K. W., Ahn, H., Shin, J., Kim, S. & Mar, W. Therapeutic effect of (Z)-3-(2,5-dimethoxyphenyl)-2-(4-methoxyphenyl) acrylonitrile (DMMA) against Staphylococcus aureus infection in a murine model. *Biochem Biophys Res Commun* **396**, 440-444, (2010).

120 Ton-That, H., Liu, G., Mazmanian, S. K., Faull, K. F. & Schneewind, O. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif. *Proc Natl Acad Sci USA* **96**, 12424-12429, (1999).

121 Kang, S. S., Kim, J. G., Lee, T. H. & Oh, K. B. Flavonols inhibit sortases and sortase-mediated Staphylococcus aureus clumping to fibrinogen. *Biol Pharm Bull* **29**, 1751-1755, (2006).

122 Kim, S. H., Shin, D. S., Oh, M. N., Chung, S. C., Lee, J. S., Chang, I. M. & Oh, K. B. Inhibition of sortase, a bacterial surface protein anchoring transpeptidase, by beta-sitosterol-3-O-glucopyranoside from Fritillaria verticillata. *Biosci Biotechnol Biochem* **67**, 2477-2479, (2003).

123 Kim, S. H., Shin, D. S., Oh, M. N., Chung, S. C., Lee, J. S. & Oh, K. B. Inhibition of the bacterial surface protein anchoring transpeptidase sortase by isoquinoline alkaloids. *Biosci Biotechnol Biochem* **68**, 421-424, (2004).

124 Oh, K. B., Oh, M. N., Kim, J. G., Shin, D. S. & Shin, J. Inhibition of sortase-mediated Staphylococcus aureus adhesion to fibronectin via fibronectin-binding protein by sortase inhibitors. *Appl Microbiol Biotechnol* **70**, 102-106, (2006).

125 Frankel, B. A., Bentley, M., Kruger, R. G. & McCafferty, D. G. Vinyl sulfones: inhibitors of SrtA, a transpeptidase required for cell wall protein anchoring and virulence in Staphylococcus aureus. *J Am Chem Soc* **126**, 3404-3405, (2004).

126 Maresso, A. W., Wu, R., Kern, J. W., Zhang, R., Janik, D., Missiakas, D. M., Duban, M. E., Joachimiak, A. & Schneewind, O. Activation of inhibitors by sortase triggers irreversible modification of the active site. *J Biol Chem* **282**, 23129-23139, (2007).

127 Scott, C. J., McDowell, A., Martin, S. L., Lynas, J. F., Vandenbroeck, K. & Walker, B. Irreversible inhibition of the bacterial cysteine protease-transpeptidase sortase (SrtA) by substrate-derived affinity labels. *Biochem J* **366**, 953-958, (2002).

128 Kruger, R. G., Barkallah, S., Frankel, B. A. & McCafferty, D. G. Inhibition of the Staphylococcus aureus sortase transpeptidase SrtA by phosphinic peptidomimetics. *Bioorg Med Chem* **12**, 3723-3729, (2004).

129 Katsara, M., Tselios, T., Deraos, S., Deraos, G., Matsoukas, M.-T., Lazoura, E., Matsoukas, J. & Apostolopoulos, V. Round and Round we Go: Cyclic Peptides in Disease. *Curr Med Chem* **13**, 2221-2232, (2006).

130 Uchiyama, F., Tanaka, Y., Minari, Y. & Tokui, N. Designing scaffolds of peptides for phage display libraries. *J Biosci Bioeng* **99**, 448-456, (2005).

131 Lee, C. V., Liang, W. C., Dennis, M. S., Eigenbrot, C., Sidhu, S. S. & Fuh, G. High-affinity human antibodies from phage-displayed synthetic fab libraries with a single framework scaffold. *J Mol Biol* **340**, 1073-1093, (2004).

132 Hackel, B. J., Kapila, A. & Wittrup, K. D. Picomolar affinity fibronectin domains engineered utilizing loop length diversity, recursive mutagenesis, and loop shuffling. *J Mol Biol* **381**, 1238-1252, (2008).

133 Koide, A., Gilbreth, R. N., Esaki, K., Tereshko, V. & Koide, S. High-affinity single-domain binding proteins with a binary-code interface. *Proc Natl Acad Sci USA* **104**, 6632-6637, (2007).

134 Bonnycastle, L. L. C., Mehroke, J. S., Rashed, M., Gong, X. & Scott, J. K. Probing the basis of antibody reactivity with a panel of constrained peptide libraries displayed by filamentous phage. *J Mol Biol* **258**, 747-762, (1996).

135 Koivunen, E., Wang, B. & Ruoslahti, E. Phage Libraries Displaying Cyclic Peptides with Different Ring Sizes: Ligand Specificities of the RGD-Directed Integrins. *Nat Biotechnol* **13**, 265-270, (1995).

136 Smith, G. P., Patel, S. U., Windass, J. D., Thornton, J. M., Winter, G. & Griffiths, A. D. Small binding proteins selected from a combinatorial repertoire of knottins displayed on phage. *J Mol Biol* **277**, 317-332, (1998).

137 Chang, H. J., Hsu, H. J., Chang, C. F., Peng, H. P., Sun, Y. K., Yu, H. M., Shih, H. C., Song, C. Y., Lin, Y. T., Chen, C. C., Wang, C. H. & Yang, A. S. Molecular evolution of cystine-stabilized miniproteins as stable proteinaceous binders. *Structure* **17**, 620-631, (2009).

138 Zoller, F., Haberkorn, U. & Mier, W. Miniproteins as phage display-scaffolds for clinical applications. *Molecules* **16**, 2467-2485, (2011).

139 Phan, T., Nguyen, H. D., Goksel, H., Mocklinghoff, S. & Brunsveld, L. Phage display selection of miniprotein binders of the Estrogen Receptor. *Chem Commun (Camb)* **46**, 8207-8209, (2010).

140 Cheng, Y.-c. & Prusoff, W. H. Relationship between the inhibition constant (Ki) and the concentration of inhibitor which causes 50 percent inhibition (IC50) of an enzymatic reaction. *Biochem Pharm* **22**, 3099-3108, (1973).

141 Bradbury, A. R., Sidhu, S., Dubel, S. & McCafferty, J. Beyond natural antibodies: the power of in vitro display technologies. *Nat Biotechnol* **29**, 245-254, (2011).

142 Rothe, A., Hosse, R. J. & Power, B. E. In vitro display technologies reveal novel biopharmaceutics. *FASEB J* **20**, 1599-1610, (2006).

143 Buller, F., Steiner, M., Scheuermann, J., Mannocci, L., Nissen, I., Kohler, M., Beisel, C. & Neri, D. High-throughput sequencing for the identification of binding molecules from DNA-encoded chemical libraries. *Bioorg Med Chem Lett* **20**, 4188-4192, (2010).

144 Glanville, J., Zhai, W., Berka, J., Telman, D., Huerta, G., Mehta, G. R., Ni, I., Mei, L., Sundar, P. D., Day, G. M., Cox, D., Rajpal, A. & Pons, J. Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proc Natl Acad Sci USA* **106**, 20216-20221, (2009).

145 Ernst, A., Gfeller, D., Kan, Z., Seshagiri, S., Kim, P. M., Bader, G. D. & Sidhu, S. S. Coevolution of PDZ domain-ligand interactions analyzed by high-throughput phage display and deep sequencing. *Mol Biosyst* **6**, 1782-1790, (2010).

146 McLaughlin, M. E. & Sidhu, S. S. Engineering and analysis of peptide-recognition domain specificities by phage display and deep sequencing. *Methods Enzymol* **523**, 327-349, (2013).

147 Dias-Neto, E., Nunes, D. N., Giordano, R. J., Sun, J., Botz, G. H., Yang, K., Setubal, J. C., Pasqualini, R. & Arap, W. Next-Generation Phage Display: Integrating and Comparing Available Molecular Tools to Enable Cost-Effective High-Throughput Analysis. *PLoS One* **4**, e8338, (2009).

148 Matochko, W. L. & Derda, R. Error analysis of deep sequencing of phage libraries: peptides censored in sequencing. *Comput Math Methods Med* **2013**, Article ID. 491612, (2013).

149 Ngubane, N. A., Gresh, L., Ioerger, T. R., Sacchettini, J. C., Zhang, Y. J., Rubin, E. J., Pym, A. & Khati, M. High-throughput sequencing enhanced phage display identifies peptides that bind mycobacteria. *PLoS One* **8**, e77844, (2013).

150 Ryvkin, A., Ashkenazy, H., Smelyanski, L., Kaplan, G., Penn, O., Weiss-Ottolenghi, Y., Privman, E., Ngam, P. B., Woodward, J. E., May, G. D., Bell, C., Pupko, T. & Gershoni, J. M. Deep Panning: steps towards probing the IgOme. *PLoS One* **7**, e41469, (2012).

151 Kim, T., Tyndel, M. S., Huang, H., Sidhu, S. S., Bader, G. D., Gfeller, D. & Kim, P. M. MUSI: an integrated system for identifying multiple specificity from very large peptide or nucleic acid data sets. *Nucleic Acids Res* **40**, e47, (2012).

152 Andreatta, M., Lund, O. & Nielsen, M. Simultaneous alignment and clustering of peptide data using a Gibbs sampling approach. *Bioinformatics* **29**, 8-14, (2013).

153 Rentero Rebollo, I. & Heinis, C. Phage selection of bicyclic peptides. *Methods* **60**, 46-54, (2013).

154 Rentero Rebollo, I., Angelini, A. & Heinis, C. Phage display libraries of differently sized bicyclic peptides. *MedChemComm* **4**, 145-150, (2013).

155 Cock, P. J., Fields, C. J., Goto, N., Heuer, M. L. & Rice, P. M. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res* **38**, 1767-1771, (2010).

156 Bragg, L. M., Stone, G., Butler, M. K., Hugenholtz, P. & Tyson, G. W. Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. *PLoS Comput Biol* **9**, e1003031, (2013).

157 Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P. & Gu, Y. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**, (2012).

158 Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J. & Pallen, M. J. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* **30**, 434-439, (2012).

159 Derda, R., Tang, S. K. Y., Li, S. C., Ng, S., Matochko, W. & Jafari, M. R. Diversity of phage-displayed libraries of peptides during panning and amplification. *Molecules* **16**, 1776-1803, (2011).

160 Rodi, D. J., Soares, A. S. & Makowski, L. Quantitative Assessment of Peptide Sequence Diversity in M13 Combinatorial Peptide Phage Display Libraries. *J Mol Biol* **322**, 1039-1052, (2002).

161 Liu, C. C., Mack, A. V., Tsao, M. L., Mills, J. H., Lee, H. S., Choe, H., Farzan, M., Schultz, P. G. & Smider, V. V. Protein evolution with an expanded genetic code. *Proc Natl Acad Sci USA* **105**, 17688-17693, (2008).

162 Ton-That, H., Liu, G., Mazmanian, S. K., Faull, K. F. & Schneewind, O. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif. *Proc Natl Acad Sci USA* **96**, 12424-12429, (1999).

163 Foster, T. J., Geoghegan, J. A., Ganesh, V. K. & Hook, M. Adhesion, invasion and evasion: the many functions of the surface proteins of Staphylococcus aureus. *Nat Rev Microbiol* **12**, 49-62, (2014).

164 Maresso, A. W. & Schneewind, O. Sortase as a target of anti-infective therapy. *Pharmacol Rev* **60**, 128-141, (2008).

165 Foster, T. J. & McDevitt, D. Surface-associated proteins of Staphylococcus aureus: their possible roles in virulence. *FEMS Microbiol Lett* **118**, 199-205, (1994).

166 Chen, F. G., Liu, B. R., Wang, D. C., Wang, L., Deng, X. M., Bi, C. W., Xiong, Y., Wu, Q. C., Cui, Y. W., Zhang, Y., Li, X. L., Wang, Y., Liu, B. & Cao, Y. G. Role of sortase A in the pathogenesis of Staphylococcus aureus-induced mastitis in mice. *FEMS Microbiol Lett* **351**, 95-103, (2014).

167 Bierne, H., Mazmanian, S. K., Trost, M., Pucciarelli, M. G., Liu, G., Dehoux, P., Jansch, L., Garcia-del Portillo, F., Schneewind, O. & Cossart, P. Inactivation of the srtA gene in Listeria monocytogenes inhibits anchoring of surface proteins and affects virulence. *Mol Microbiol* **43**, 869-881, (2002).

168 Bolken, T. C., Franke, C. A., Jones, K. F., Zeller, G. O., Jones, C. H., Dutton, E. K. & Hruby, D. E. Inactivation of the srtA gene in Streptococcus gordonii inhibits cell wall anchoring of surface proteins and decreases in vitro and in vivo adhesion. *Infect Immun* **69**, 75-80, (2001).

169 Fitzgerald-Hughes, D., Devocelle, M. & Humphreys, H. Beyond conventional antibiotics for the future treatment of methicillin-resistant Staphylococcus aureus infections: two novel alternatives. *FEMS Immunol Med Microbiol* **65**, 399-412, (2012).

170 Hofmann, F. T., Szostak, J. W. & Seebeck, F. P. In Vitro Selection of Functional Lantipeptides. *J Am Chem Soc* **134**, 8038-8041, (2012).

171 Chen, S., Touati, J. & Heinis, C. Tracking chemical reactions on the surface of filamentous phage using mass spectrometry. *Chem Commun* **50**, 5267-5269, (2014).

172 Suree, N., Liew, C. K., Villareal, V. A., Thieu, W., Fadeev, E. A., Clemens, J. J., Jung, M. E. & Clubb, R. T. The structure of the Staphylococcus aureus sortase-substrate complex reveals how the universally conserved LPXTG sorting signal is recognized. *J Biol Chem* **284**, 24465-24477, (2009).

173 Ythier, M., Entenza, J. M., Bille, J., Vandenesch, F., Bes, M., Moreillon, P. & Sakwinska, O. Natural variability of in vitro adherence to fibrinogen and fibronectin does not correlate with in vivo infectivity of Staphylococcus aureus. *Infect Immun* **78**, 1711-1716, (2010).

174 Hu, P., Huang, P. & Chen, W. M. Curcumin inhibits the Sortase A activity of the Streptococcus mutans UA159. *Appl Biochem Biotechnol* **171**, 396-402, (2013).

175 Hu, P., Huang, P. & Chen, M. W. Curcumin reduces Streptococcus mutans biofilm formation by inhibiting sortase A activity. *Arch Oral Biol* **58**, 1343-1348, (2013).

176 Huang, P., Hu, P., Zhou, S. Y., Li, Q. & Chen, W. M. Morin Inhibits Sortase A and Subsequent Biofilm Formation in Streptococcus mutans. *Curr Microbiol* **68**, 47-52, (2014).

177 Grundmeier, M., Hussain, M., Becker, P., Heilmann, C., Peters, G. & Sinha, B. Truncation of fibronectin-binding proteins in Staphylococcus aureus strain Newman leads to deficient adherence and host cell invasion due to loss of the cell wall anchor function. *Infect Immun* **72**, 7155-7163, (2004).

178 Nelson, J. W., Chamessian, A. G., McEnaney, P. J., Murelli, R. P., Kazmiercak, B. I. & Spiegel, D. A. A Biosynthetic Strategy for Re-engineering the Staphylococcus aureus Cell Wall with Non-native Small Molecules. *ACS Chem Biol* **5**, 1147-1155, (2010).

179 Manaskova, S. H., Nazmi, K., van Belkum, A., Bikker, F. J., van Wamel, W. J. B. & Veerman, E. C. I. Synthetic LPETG-Containing Peptide Incorporation in the Staphylococcus aureus Cell-Wall in a Sortase A- and Growth Phase-Dependent Manner. *PLoS One* **9**, (2014).

180 DeDent, A., Bae, T., Missiakas, D. M. & Schneewind, O. Signal peptides direct surface proteins to two distinct envelope locations of Staphylococcus aureus. *EMBO J* **27**, 2656-2668, (2008).

181 Hoogenboom, H. R., Griffiths, A. D., Johnson, K. S., Chiswell, D. J., Hudson, P. & Winter, G. Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains. *Nucleic Acids Res* **19**, 4133-4137, (1991).

Curriculum Vitae

Inmaculada RENTERO REBOLLO

born 31 August 1987 in Madrid, Spain

Ecole Polytechnique Fédérale de Lausanne (EPFL)
BCH 5201 · 1015 Lausanne · Switzerland
inmaculada.rentero@epfl.ch

Education

08/2010 – present	Ph.D. in Chemical Biology Swiss Federal Institute of Technology (EPFL), Lausanne Supervisor: Prof. Christian Heinis
09/2008 – 07/2010	*Licenciatura* (Licence) in Biotechnology Autonomous University of Barcelona (UAB), Spain Research project supervisor: Dr. Efren Riu Undergraduate internships: - Max Planck Institute for Biochemistry, Munich, Germany - Molecular Biology Institute in Barcelona (IBMB), Spain - Biomedical Research Institue (IRB), Barcelona, Spain
09/2005 – 07/2008	*Primer ciclo* (Bachelor) in Biology University of Alcalá de Henares (UAH), Spain

Fellowships and awards

- National Award for Excellence in Academic Performance (3rd prize), Spanish Government, 2012

- Excellence Graduation Award (Biotechnology graduates, top 2%), Autonomous University of Barcelona, 2010
- End-of-studies Collaboration Scholarship, Spanish Government, academic year 2009/10
- Amgen Scholars Summer Internship, Amgen Foundation, 2009
- JAE-Intro Summer Scholarship, Spanish National Council for Research (CSIC), 2009
- JAE-Intro Summer Scholarship, Spanish National Council for Research (CSIC), 2008
- Madrid Regional Government Excellence Awards (academic years 2005/06, 2006/07, 2007/08)

Publications

- I. Rentero Rebollo, M. Sabisz, V. Baeriswyl, C. Heinis. Identification of target-binding peptide motifs by high-throughput sequencing of phage-selected peptides. *Nucleic Acids Res.*, 2014.
- S. Belloto, S. Chen, I. Rentero Rebollo, H. Wegner, C. Heinis. Phage selection of photoswitchable peptide ligands. *J. Am. Chem. Soc,* 2014, **136**:5880-3.
- S. Chen, I. Rentero Rebollo, S. Buth, J. Morales-Sanfrutos, J. Touati, P. Leiman, C. Heinis. Bicyclic peptide ligands pulled out of cysteine-rich peptide libraries. *J. Am. Chem. Soc.*, 2013, **135**:6562-9.
- I. Rentero Rebollo, C. Heinis. Phage selection of bicyclic peptides. *Methods,* 2013, **60**:46-54.
- I. Rentero Rebollo, A. Angelini, C. Heinis. Phage Display Libraries of Differently Sized Bicyclic Peptides. *Med. Chem. Commun.,* 2013, **4**:145-50.
- N. Salvi, R. Buratto, A. Bornet, S. Ulzega, I. Rentero Rebollo, A. Angelini, C. Heinis, G. Bodenhausen. Boosting Sensitivity of Ligand-Protein Screening by NMR of Long-Lived States. *J. Am. Chem. Soc.,* 2012, **134**(27):11076-9.
- I. Rentero Rebollo, C. Heinis. Screening of large molecule diversities by phage display. *Chimia*, 2011, **65**(11):832-5